frontiers | Research Topics

# Fisheries and aquaculture genetics

**Edited by**
Yuzine Esa, Siti Nor, Md Samsul Alam and Nguyen Hong Nguyen

**Published in**
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-2888-4
DOI 10.3389/978-2-8325-2888-4

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Fisheries and aquaculture genetics

**Topic editors**

Yuzine Esa — Putra Malaysia University, Malaysia
Siti Nor — University of Malaysia Terengganu, Malaysia
Md Samsul Alam — Bangladesh Agricultural University, Bangladesh
Nguyen Hong Nguyen — University of the Sunshine Coast, Australia

**Citation**

Esa, Y., Nor, S., Alam, M. S., Nguyen, N. H., eds. (2023). *Fisheries and aquaculture genetics*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2888-4

# Table of contents

frontiers | Frontiers in Genetics

DATA REPORT
published: 17 June 2022
doi: 10.3389/fgene.2022.911446

# First Report of Chromosome-Level Genome Assembly for Flathead Grey Mullet, *Mugil cephalus* (Linnaeus, 1758)

*Mudagandur S. Shekhar*[1], *Vinaya Kumar Katneni*[1]*, *Ashok Kumar Jangam*[1], *Karthic Krishnan*[1], *Sudheesh K. Prabhudas*[1], *Jesudhas Raymond Jani Angel*[1], *Krishna Sukumaran*[2], *Muniyandi Kailasam*[2] and *Joykrushna Jena*[3]

[1]Nutrition Genetics and Biotechnology Division, Indian Council of Agricultural Research-Central Institute of Brackishwater Aquaculture, Chennai, India, [2]Finfish Culture Division, ICAR-Central Institute of Brackishwater Aquaculture, Chennai, India, [3]Indian Council of Agricultural Research, New Delhi, India

**Keywords: whole genome, repeat analysis, synteny, grey mullet, PacBio Sequel II, genome annotation**

OPEN ACCESS

***Edited by:***
*Xu Wang,*
*Auburn University, United States*

***Reviewed by:***
*Eric M. Hallerman,*
*Virginia Tech, United States*
*Xiaoli Ma,*
*Jiangsu Normal University, China*

****Correspondence:***
*Vinaya Kumar Katneni*
*Vinaya.Katneni@icar.gov.in*

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

***Received:*** *02 April 2022*
***Accepted:*** *06 May 2022*
***Published:*** *17 June 2022*

***Citation:***
*Shekhar MS, Katneni VK, Jangam AK, Krishnan K, Prabhudas SK, Jani Angel JR, Sukumaran K, Kailasam M and Jena J (2022) First Report of Chromosome-Level Genome Assembly for Flathead Grey Mullet, Mugil cephalus (Linnaeus, 1758). Front. Genet. 13:911446. doi: 10.3389/fgene.2022.911446*

## INTRODUCTION

Amid dwindling production from capture fisheries, aquaculture is an important source of high-quality protein (FAO Food and Agriculture Organization of the United Nations, 2020). In addition, aquaculture generates employment and contributes to eradicating poverty. *Mugil cephalus*, flathead grey mullet, is globally distributed and inhabits inshore seas, estuaries, and brackish water areas (Whitfield et al., 2012). It is of important commercial value to global fisheries and aquaculture with high demand for mullet roe. This species belongs to the family Mugilidae which comprises 26 genera and 80 valid species (Fricke et al., 2022). The adaptability of *M. cephalus* to varied aquatic environments at different life stages and tolerance to a wide range of salinities and temperatures occurring in tropical, subtropical, and temperate coastal waters make it an important cultivatable fish species across the world. The whole-genome sequence information for an aquaculture species has potential applications in genomic selection and selective breeding for sustainable production and improvement of desirable traits, such as disease resistance and growth.

Within the family Mugilidae, a draft genome assembly was first reported for redlip mullet, *Liza haematocheila* (0.74 Gb), which has 1,453 contigs with an N50 length of 3.9 Mb (Liyanage et al., 2019). Later, chromosome-level assembly of this fish was generated with Oxford Nanopore long-read, single-tube long fragment reads (stLFR), and HiC chromatin interaction data, which are 652.91 Mb length in 514 scaffolds with contig and scaffold N50 lengths of 7.21 and 28.01 Mb, respectively (Zhao et al., 2021). For *M. cephalus*, no whole-genome assembly with long-read data is available. Previous reports have attempted genome assembly of this fish at a very low sequence depth using Illumina technology (Dor et al., 2016, 2020). Thus, in the absence of a reference genome for *M. cephalus*, this study aimed to decipher the whole-genome sequence, which will provide baseline information needed to implement genetic improvement programs. The integration of genome information into fisheries and aquaculture management is important to ensure long-term sustainable fishery harvest and aquaculture production. In the present study, a combination of PacBio, Illumina, and Arima Hi-C technologies were applied to construct the genome assembly of *M. cephalus*, an economically important brackish water aquaculture species.

## Value of Data

– The whole-genome sequence assembly generated for *M. cephalus* can be used as a reference genome for the family Mugilidae.
– The high-quality, chromosome-level genome assembly along with the predicted protein sequences would help gain insights into desirable traits through gene expression studies.

**TABLE 1 |** Summary of *Mugil cephalus* genome assembly and annotation.

| Features | Statistics |
|---|---|
| No. of contigs | 848 |
| Contig N50 size (Mbp) | 20.15 |
| Contig L50 | 14 |
| Contig total length (Mbp) | 643.89 |
| No. of scaffolds | 583 |
| Scaffold N50 size (Mbp) | 28.32 |
| Scaffold L50 | 10 |
| Total scaffold length (Mbp) | 644.11 |
| Number of pseudo-chromosomes | 24 |
| Total length of pseudo-chromosomes (Mbp) | 634.85 |
| GC % | 41.9 |
| No. of protein coding genes | 27,269 |

– The whole-genome assembly provides baseline information needed to implement genetic improvement programs for this commercially important species.

# MATERIALS AND METHODS

## Specimen of *M. cephalus*

A specimen of *M. cephalus* maintained at the Muthukadu Experimental Station of ICAR—CIBA (Chennai, India) was used for generating the sequence data required for genome assembly. The species identity of the specimen was confirmed based on the partial sequence of the barcode gene, cytochrome C oxidase I (COI). Briefly, genomic DNA was isolated using the conventional CTAB method (Mirimin and Roodt-Wilding, 2015) from muscle tissue, and its concentration and quality were assessed using a nanodrop 2000C (Thermo Scientific, Waltham, Massachusetts, United States). A partial fragment of the COI gene was amplified using high-fidelity 2X PCR Master Mix (New England Biolabs, Ipswich, Massachusetts, United States) with the primers, F2 - 5′TCGACTAATCACAAAGAC ATCGGCAC3′ and R1 - 5′TAGACTTCTGGGTGGCCAAAG AATCA3′ (Ward et al., 2005). The amplification conditions were initial denaturation at 98°C for 30 s; 32 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s; followed by a final extension at 72°C for 2 min. The PCR product of 707 bp was gel-extracted using a QIAquick gel extraction kit (Qiagen, Hilden, Germany) and sequenced bidirectionally using an ABI 3730 sequencer (Applied Biosystems, Waltham, Massachusetts, United States). The partial sequence of the COI gene was submitted to GenBank with accession number, MW584357. The sequence of MW584357 was subjected to phylogenetic analysis along with 476 accessions of the genus *Mugil* (**Supplementary Table S1**) sourced from the Barcode of Life Data systems database[1] and accession of *Chelon labrosus* as an outgroup. Initially, all the 478 sequences were aligned in BioEdit version 7.2.5 (Hall, 1999) to generate a consensus alignment (516 bp) which was used to build a Maximum Likelihood tree in RAxML version 8.2.9 (Stamatakis, 2014) with the GTRGAMMAI model and a random seed value of 12,345. The final tree was visualized using FigTree v1.4.3[2], which revealed the clustering of sequence MW584357 with other accessions of *M. cephalus* (**Supplementary Figure S1**).

## PacBio Sequel II Data Generation

Genomic DNA was extracted from the muscle tissue of *M. cephalus* using the blood and cell culture DNA midi kit (Qiagen). DNA quantification was carried out using a Qubit 4.0 fluorometer (Thermofisher Scientific). Library preparation was performed with the SMRTbell® Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, California, United States), and size selection was carried out using BluePippin™ (Sage Science, Beverly, Massachusetts, United States). The library was sequenced on the PacBio Sequel II platform (Pacific Biosciences) to generate the sequence data. About 257.7 Gb of sequence data in 15,028,480 subreads was generated with a subread N50 of 28,748 bp (**Supplementary Table S2**).

## Illumina Data Generation

Genomic DNA extracted from the muscle tissue of *M. cephalus* was used for Illumina library preparation with an insert size of 200–300 bp using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (New England Biolabs). The PCR products used for the construction of the library were purified with the AMPure XP reagent (Beckman Coulter, Brea, California, United States). The library was checked for size distribution by Agilent 2,100 Bioanalyzer (Agilent Technologies, Santa Clara, California, United States) and sequenced on Illumina NovaSeq 6,000, S4 Flow Cell (2 × 150 bp read length). About 640 million reads/96.1 Gb data were generated with 93% of Q30 bases (**Supplementary Table S3**). The Illumina paired-end reads were used to assess genome length and also to correct the erroneous bases in genome assembly contigs.

## HiC Data Generation

For HiC data generation, tissue crosslinking and proximity ligation were performed using the Phase Genomics Kit (Phase Genomics, Seattle, Washington, United States) followed by Illumina compatible sequencing library preparation. The library was sequenced on the Illumina NovaSeq6000 platform in a 150-bp paired-end mode to generate 181 million paired reads (54.31 Gb), of which 90.74% (49.28 Gb) of bases had a Q30 quality score (**Supplementary Table S4**). The HiC reads were used in the scaffolding of assembly contigs.

## Genome Length Assessment

The *k*-mer frequency for the grey mullet genome was estimated from the Illumina paired-end reads using Jellyfish 2.2.3 (Marçais and Kingsford, 2011) to count the canonical 21 *k*-mers with the hash size as 20G. The count histogram was later provided as the input to the online tool Genomescope (Vurture et al., 2017) to estimate genome haploid length, heterozygosity, and repeat content. The *k*-mer analysis revealed the genome haploid length to be 594 Mb and genome repeat length as 48 Mb

[1]www.boldsystems.org

[2]http://tree.bio.ed.ac.uk/software/figtree/

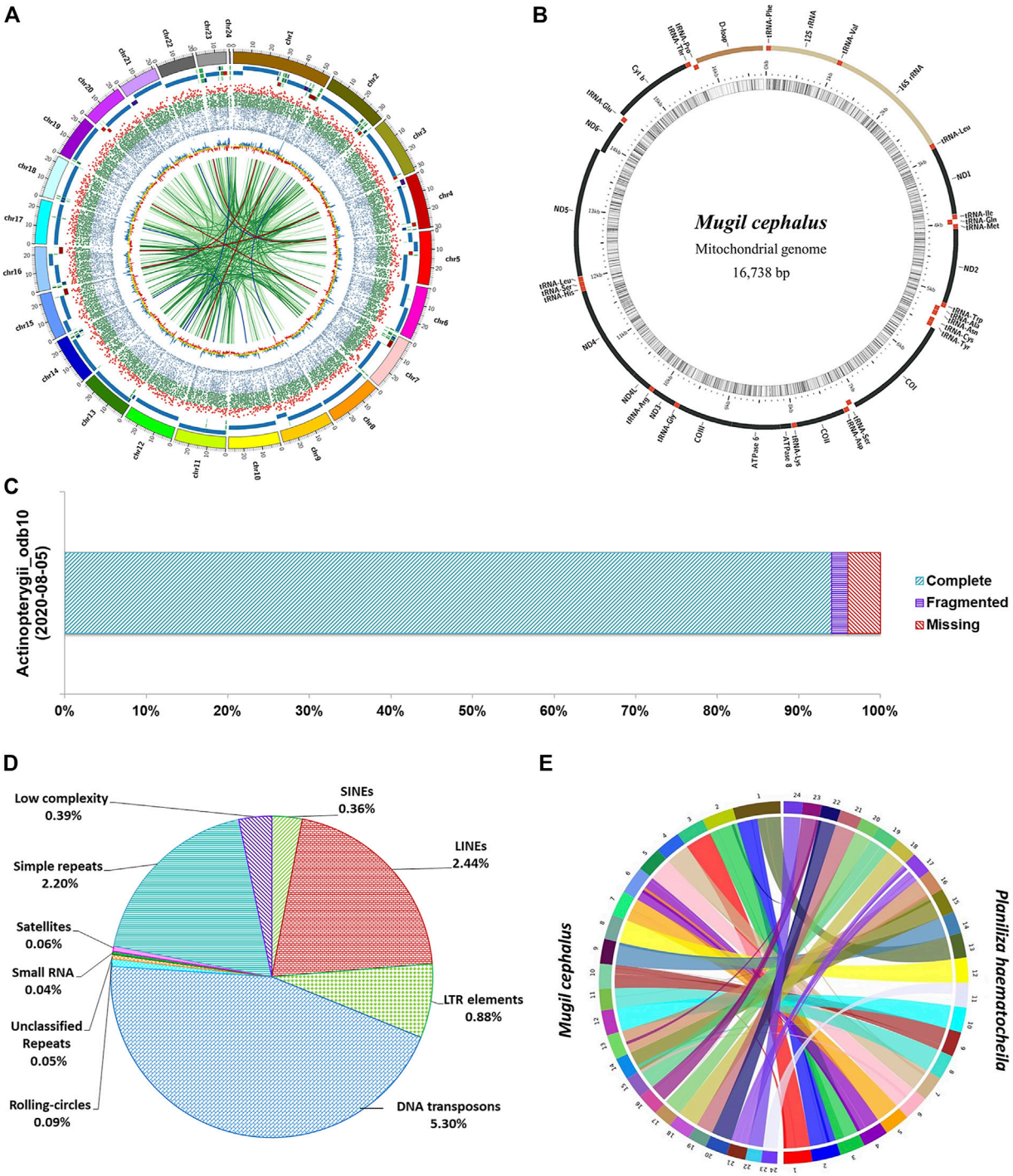



**FIGURE 1 |** Assembly of the *M. cephalus* genome. **(A)** Circos plot depicting the 24 pseudo-chromosomes in the *M. cephalus* assembly and their features. The track 1 (outermost) depicts 24 pseudo-chromosomes; track 2 depicts contigs represented as tiles (contigs <5 Mb are shown in variable colors); track 3 shows protein-coding genes as a scatter plot (gene length of <5 kb in blue, 5–30 kb in green >30 kb in red); track 4 shows GC content of the genome as a yellow line plot (GC values >43 in blue and <41 in red); track 5 (innermost) depicts self synteny in the genome shown as a link plot (alignment lengths over 1,500 bp in light green, 2–3 kb in dark green, 3–3.5 kb in dark blue and >3.5 kb in dark red). **(B)** Map of the complete mitochondrial DNA genome of *M. cephalus*. **(C)** BUSCO scores illustrating the completeness of the *M. cephalus* genome assembly. **(D)** Profile of repetitive elements in the *M. cephalus* genome assembly. **(E)** Synteny plot between redlip mullet and grey mullet genome assemblies.

(**Supplementary Figure S2**). The genome size of *M. cephalus* was reported to be 857 Mb based on the flow cytometry principle (Raymond et al., 2022).

## Genome Assembly

Genome assembly was performed with the WTDBG2.5 (Ruan and Li, 2020) tool by limiting the usage of PacBio subreads to >30 kb length covering 100 Gb of data. The *de novo* contig-level assembly contained 848 contigs with an N50 length of 20.15 Mb. Polishing of these contigs for base errors and indels with Illumina short reads using POLCA (Zimin and Salzberg, 2020) brought the base consensus quality to 99.99% (**Supplementary Table S5**). Scaffolding was performed on the polished assembly by using juicer and 3D-DNA scripts from the Genome Assembly Cookbook (Dudchenko et al., 2017). Initially, site positions were generated based on the DpnII enzyme used for generating 181 million HiC read pairs Later, a juicer script was used to generate contact maps in 3D space, and a 3D-DNA script was used to anchor fragments to pseudo-chromosomes which were visualized in Juicebox v1.11.08 (**Supplementary Figure S3**).

At the scaffold level, the assembly is 644 Mb in length in 583 scaffolds with an N50 of 28.32 Mb. The *M. cephlus* karyotype is reported to contain 24 chromosome pairs (Gornung et al., 2001). Accordingly, the longest 24 scaffolds represented 98.56% (634 Mb) of the scaffold-level assembly length and hence are designated as pseudo-chromosomes (**Table 1**). The mitochondrial genome of *Mugil cephalus* was obtained as a single scaffold of 16,738 bp in the assembly. The mitochondrial genome was annotated (**Figure 1B**) using the MitoAnnotator tool[3] (Iwasaki et al., 2013). The genome assembly presented for *M. cephalus* in this study is superior to the assembly reported by Dor et al. (2016) with 480,389 scaffolds and to that reported by Dor et al. (2020) with 4,505 scaffolds in terms of N50 statistics and average scaffold length. The genome assembly reported in this study is of a shorter length than the estimated genome length reported in previous studies (Hinegardner and Rosen, 1972; Dor et al., 2016, 2020; Raymond et al., 2022). The genome completeness for *M. cephalus* assembly was assessed to be 96% complete (**Figure 1C**) using BUSCO v5.2.2 (Seppey et al., 2019) against the Actinopterygii_odb10 (2020-08-05) orthologous dataset (Kriventseva et al., 2019).

## Repeat Prediction

Repeat masking was performed by using the Repeat masker v4.0.9[4] available in the genome analysis module of Omicsbox v2.0.36 (Bioinformatics, 2019). The RMBlast search was performed for 24 pseudo-chromosomes with a total of 634,849,760 bp against the Eukaryota subset of the RepBase Repeat Masker Edition-20181026[5]. The interspersed, low-complexity, and simple-sequence repeats were masked to obtain the complete repeat profile of *M. cephalus* genome assembly. The total repetitive elements in *M. cephalus* assembly constitute 11.72% (74,376,509 bp) of assembly length as shown in **Supplementary Table S6**. DNA transposons (5.3%, 33,664,198 bp) and the satellite repeats (0.06%, 362,353 bp) contributed the highest and the lowest, respectively, to the total repeat content of the genome (**Figure 1D**). The LINEs were the major contributor for retroelements, followed by LTR elements and SINEs accounting for 2.44, 0.88, and 0.36%, respectively. Simple-sequence repeats and low-complexity repeats covered 2.2% (13,966,215 bp) and 0.39% (2,504,867 bp) of the genome, respectively. The MISA v1.0 tool was used to generate the simple sequence repeat profile containing a total of 491,676 SSR elements in assembly (**Supplementary Figure S4**).

## Gene Prediction

The pseudo-chromosomes soft-masked for interspersed sequence repeats avoiding the low-complexity and simple-sequence repeats were used for gene prediction. Structural annotation of the *Mugil cephalus* genome was performed by following the approach as described previously (Katneni et al., 2022), with minor modifications. A combination of *ab initio* and evidence-based approaches was used to obtain the final protein-coding gene set. Briefly, *ab initio* prediction was performed using Augustus v3.3.3 (Stanke et al., 2006) and GeneMark-ES v4.59 (Lomsadze et al., 2005) on the repeat-masked genome. The Illumina paired-end RNA-seq data of fish tissues and developmental stages accessed from GenBank (SRR13039561-66; SRR15243984-88 and SRR16311547-49) were mapped to the genome assembly using Hisat2 v2.2.0 (Kim et al., 2019) aligner. Then, StringTie v2.1.4 (Pertea et al., 2015) was used on this sorted alignment file to assemble the transcripts which were processed in TransDecoder v5.5.0[6] to obtain the Open-Reading Frames (ORFs). Similarly, PacBio IsoSequence reads (NCBI Bioproject: PRJNA675305) were aligned to the genome assembly using GMAP v2020-06-30 (Wu and Watanabe, 2005) to generate valid gene structures using PASA v2.4.1 (Haas et al., 2003) and then finally to generate genome coordinates using TransDecoder v5.5.0[6]. Furthermore, evidence was gathered from the proteins of related species (*Archocentrus centrachus*, *Astatotilapia calliptera*, *Gambusia affinis*, *Melanotaenia boesemani*, *Maylandia zebra*, *Oreochromis niloticus*, *Oryzias melastigma*, *Parambassis ranga*, *Salarias fasciatus*, and *Xiphophorus maculatus* genome assemblies) by aligning them to the *M. cephalus* genome using GenomeThreader v1.7.3 (Gremme, 2012). Finally, all the evidence from *ab initio* gene predictions, Isosequence reads, RNA-seq reads, and protein data of related species were combined using Evidence Modeler (Haas et al., 2008) to obtain the consensus and non-redundant protein-coding gene set containing 30,882 genes. Then, AGAT[7] was used to filter out the incomplete gene models, resulting in the final gene set of 27,269 genes (**Supplementary Table S7**). Noncoding genes in the *M. cephalus* genome were identified by aligning the repeat-masked assembly to the Rfam[8] database using infernal v1.1.4 (Nawrocki and Eddy, 2013). A total of 5,165 non-coding RNAs were identified (**Supplementary Table S8**), out of which 5S_RNA, tRNA, and histone3 were in abundance.

[3]http://mitofish.aori.u-tokyo.ac.jp/annotation/input.html
[4]http://www.repeatmasker.org/
[5]https://www.girinst.org/repbase/
[6]https://github.com/TransDecoder
[7]https://github.com/NBISweden/AGAT
[8]http://rfam.xfam.org/

## Functional Annotation

Homology-based genome annotation was performed with the blastx tool (Altschul et al., 1990) using the Actinopterygii (txid7898) non-redundant database of GenBank. The protein domains and orthology-based annotations were executed by the Interproscan module and EggNOG mapper module (Huerta-Cepas et al., 2019), as implemented in OmicsBox v2.0.36 (Bioinformatics, 2019), respectively. Then, all the annotations were merged using OmicsBox v2.0.36 (Bioinformatics, 2019) to obtain the gene ontology of the annotated genes and map the enzyme codes. The pathway details of the annotated genes were obtained by mapping against the KEGG database (Kanehisa et al., 2017). Blastx hits were obtained for 22,334 (81.9%) transcripts from which 18,806 (69%) were functionally annotated and 8,113 (29.8%) were mapped with enzyme codes (**Supplementary Figure S5**). The transferases and hydrolases were the most dominant enzyme classes expressed (**Supplementary Figure S6**). Gene ontology analysis revealed that the most expressed GO categories were gene expression (Biological processes), metal ion binding (Molecular function), and intracellular membrane-bounded organelle (Cellular components) in *M. cephalus* proteins (**Supplementary Figure S7**).

## Synteny Analysis Between *Mugil cephalus* and *Planiliza haematocheila*

Comparative genome analysis was performed to characterize genome-wide syntenic regions between redlip mullet (Zhao et al., 2021) and grey mullet (this study) genomes. The 24 chromosome–level scaffolds of the redlip mullet assembly (648.41 Mb) obtained from the China National Gene Bank Database (CNGBdb: CNP0001604) were aligned to the 24 pseudo-chromosomes of the grey mullet assembly (634.84 Mb) using SyMAP 4.2 (Soderlund et al., 2006). A total of 93 synteny blocks were observed between the two genomes. Among them, 24 blocks of the grey mullet genome were found to align against 25 blocks of the redlip mullet genome with block lengths greater than 10 Mb (**Figure 1E**; **Supplementary Table S9**).

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA675305; https://figshare.com/, https://doi.org/10.6084/m9.figshare.19499054.v1.

## ETHICS STATEMENT

The animal study was reviewed and approved by the Institute Animal Ethics Committee (CIBA/IAEC/2021-25) of ICAR—Central Institute of Brackishwater Aquaculture, Chennai, India.

## AUTHOR CONTRIBUTIONS

MS and JoJ, designed and conducted the study. VK, AJ, KK, and SP performed the sequencing analysis and genome assembly. JeJ, KS, and MK provided the samples and conducted phylogenetic analysis. MS wrote the manuscript with inputs from all authors. All the authors read and approved the final version of the manuscript.

## FUNDING

This work was supported by ICAR-Consortia Research Platform on Genomics, Indian Council of Agricultural Research, New Delhi, India.

## ACKNOWLEDGMENTS

The authors acknowledge the Director, ICAR-CIBA for providing the infrastructure and support required to conduct the study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.911446/full#supplementary-material

## REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Bioinformatics, B. (2019). OmicsBox-Bioinformatics Made Easy. Available at: https://www.biobam.com/omicsbox/.

Dor, L., Shirak, A., Curzon, A. Y., Rosenfeld, H., Ashkenazi, I. M., Nixon, O., et al. (2020). Preferential Mapping of Sex-Biased Differentially-Expressed Genes of Larvae to the Sex-Determining Region of Flathead Grey Mullet (*Mugil cephalus*). *Front. Genet.* 11, 839. doi:10.3389/fgene.2020.00839

Dor, L., Shirak, A., Rosenfeld, H., Ashkenazi, I. M., Band, M. R., Korol, A., et al. (2016). Identification of the Sex-Determining Region in Flathead Grey Mullet (*Mugil cephalus* ). *Anim. Genet.* 47, 698–707. doi:10.1111/age.12486

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). De Novo assembly of the *Aedes aegypti* Genome Using Hi-C Yields Chromosome-Length Scaffolds. *Science* 356, 92–95. doi:10.1126/science.aal3327

FAO Food and Agriculture Organization of the United Nations (2020). The State of World Fisheries and Aquaculture 2020. Available at: https://www.fao.org/documents/card/en/c/ca9229en/.

Fricke, R., Eschmeyer, W. N., and Fong, J. D. (2022). Eschmeyer's Catalog of Fishes: Genera/Species by Family/Subfamily. Available at: http://researcharchive.calacademy.org/research/ichthyology/catalog/SpeciesByFamily.asp (accessed on 01 Mar, 2022).

Gornung, E., Cordisco, C., Rossi, A., Innocentiis, D. S., Crosetti, D., and Sola, L. (2001). Chromosomal Evolution in Mugilidae: Karyotype Characterization of *Liza Saliens* and Comparative Localization of Major

and Minor Ribosomal Genes in the Six Mediterranean Mullets. *Mar. Biol.* 139, 55–60.

Gremme, G. (2012). *Computational Gene Structure Prediction*. Hamburg: Universität Hamburg. PhD thesis.

Haas, B. J., Delcher, A. L., Mount, S. M., Wortman, J. R., Smith, R. K., Hannick, L. I., et al. (2003). Improving the *Arabidopsis* Genome Annotation Using Maximal Transcript Alignment Assemblies. *Nucleic Acids Res.* 31, 5654–5666. doi:10.1093/nar/gkg770

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7–R22. doi:10.1186/gb-2008-9-1-r7

Hall, T. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Hinegardner, R., and Rosen, D. E. (1972). Cellular DNA Content and the Evolution of Teleostean Fishes. *Am. Nat.* 106, 621–644. doi:10.1086/282801

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., et al. (2019). eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. *Nucleic Acids Res.* 47, D309–D314. doi:10.1093/nar/gky1085

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., et al. (2013). MitoFish and MitoAnnotator: A Mitochondrial Genome Database of Fish with an Accurate and Automatic Annotation Pipeline. *Mol. Biol. Evol.* 30, 2531–2540. doi:10.1093/molbev/mst141

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* 45, D353–D361. doi:10.1093/nar/gkw1092

Katneni, V. K., Shekhar, M. S., Jangam, A. K., Krishnan, K., Prabhudas, S. K., Kaikkolante, N., et al. (2022). A Superior Contiguous Whole Genome Assembly for Shrimp (*Penaeus Indicus*). *Front. Mar. Sci.* 8, 8354. doi:10.3389/fmars.2021.808354

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat. Biotechnol.* 37, 907–915. doi:10.1038/s41587-019-0201-4

Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., et al. (2019). OrthoDB V10: Sampling the Diversity of Animal, Plant, Fungal, Protist, Bacterial and Viral Genomes for Evolutionary and Functional Annotations of Orthologs. *Nucleic Acids Res.* 47, D807–D811. doi:10.1093/nar/gky1053

Liyanage, D. S., Oh, M., Omeka, W. K. M., Wan, Q., Jin, C. N., Shin, G. H., et al. (2019). First Draft Genome Assembly of Redlip Mullet (*Liza Haematocheila*) from Family Mugilidae. *Front. Genet.* 10, 1246. doi:10.3389/fgene.2019.01246

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., and Borodovsky, M. (2005). Gene Identification in Novel Eukaryotic Genomes by Self-Training Algorithm. *Nucleic Acids Res.* 33, 6494–6506. doi:10.1093/nar/gki937

Marçais, G., and Kingsford, C. (2011). A Fast, Lock-free Approach for Efficient Parallel Counting of Occurrences of *K*-Mers. *Bioinformatics* 27, 764

Mirimin, L., and Roodt-Wilding, R. (2015). Testing and Validating a Modified CTAB DNA Extraction Method to Enable Molecular Parentage Analysis of Fertilized Eggs and Larvae of an Emerging South African Aquaculture Species, the Dusky Kob *Argyrosomus Japonicus*. *J. Fish. Biol.* 86, 1218–1223. doi:10.1111/jfb.12639

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold Faster RNA Homology Searches. *Bioinformatics* 29, 2933–2935. doi:10.1093/bioinformatics/btt509

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie Enables Improved Reconstruction of a Transcriptome from RNA-Seq Reads. *Nat. Biotechnol.* 33, 290–295. doi:10.1038/nbt.3122

Raymond, J. A. J., Shekhar, M. S., Katneni, V. K., Jangham, A. K., Prabhudas, S. K., Kaikkolante, N., et al. (2022). Comparative Genome Size Estimation of Different Life Stages of Grey Mullet, *Mugil cephalus* Linnaeus, 1758 by Flow Cytometry. *Aquac. Res.* 53, 1151–1158. doi:10.1111/are.15639

Ruan, J., and Li, H. (2020). Fast and Accurate Long-Read Assembly with Wtdbg2. *Nat. Methods* 17, 155–158. doi:10.1038/s41592-019-0669-3

Seppey, M., Manni, M., and Zdobnov, E. M. (2019). "BUSCO: Assessing Genome Assembly and Annotation Completeness," in *Gene Prediction*. Editor M. Kollmar (Humana, NY: Springer), 227–245. doi:10.1007/978-1-4939-9173-0_14

Soderlund, C., Nelson, W., Shoemaker, A., and Paterson, A. (2006). SyMAP: A System for Discovering and Viewing Syntenic Regions of FPC Maps. *Genome Res.* 16, 1159–1168. doi:10.1101/gr.5396706

Stamatakis, A. (2014). RAxML Version 8: a Tool for Phylogenetic Analysis and Post-analysis of Large Phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: *Ab Initio* Prediction of Alternative Transcripts. *Nucleic Acids Res.* 34, W435–W439. doi:10.1093/nar/gkl200

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: Fast Reference-free Genome Profiling from Short Reads. *Bioinformatics* 33, 2202–2204. doi:10.1093/bioinformatics/btx153

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., and Hebert, P. D. N. (2005). DNA Barcoding Australia's Fish Species. *Phil. Trans. R. Soc. B* 360, 1847–1857. doi:10.1098/rstb.2005.1716

Whitfield, A. K., Panfili, J., and Durand, J.-D. (2012). A Global Review of the Cosmopolitan Flathead Mullet *Mugil cephalus* Linnaeus 1758 (Teleostei: Mugilidae), with Emphasis on the Biology, Genetics, Ecology and Fisheries Aspects of This Apparent Species Complex. *Rev. Fish. Biol. Fish.* 22, 641–681. doi:10.1007/s11160-012-9263-9

Wu, T. D., and Watanabe, C. K. (2005). GMAP: A Genomic Mapping and Alignment Program for mRNA and EST Sequences. *Bioinformatics* 21, 1859–1875. doi:10.1093/bioinformatics/bti310

Zhao, N., Guo, H.-B., Jia, L., Guo, H.-B., Jia, L., Deng, Q.-X., et al. (2021). High-quality Chromosome-Level Genome Assembly of Redlip Mullet (Planiliza Haematocheila). *Zool. Res.* 42, 796–799. doi:10.24272/j.issn.2095-8137.2021.255

Zimin, A. V., and Salzberg, S. L. (2020). The Genome Polishing Tool POLCA Makes Fast and Accurate Corrections in Genome Assemblies. *PLoS Comput. Biol.* 16, e1007981–8. doi:10.1371/journal.pcbi.1007981

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

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 23 June 2022
doi: 10.3389/fgene.2022.936806

# Identification of Haplotypes Associated With Resistance to Bacterial Cold Water Disease in Rainbow Trout Using Whole-Genome Resequencing

*Sixin Liu[1]*, Kyle E. Martin[2], Guangtu Gao[1], Roseanna Long[1], Jason P. Evenhuis[1], Timothy D. Leeds[1], Gregory D. Wiens[1] and Yniv Palti[1]*

[1]National Center for Cool and Cold Water Aquaculture, Agricultural Research Service, United States Department of Agriculture, Kearneysville, WV, United States, [2]Troutlodge Inc., Sumner, WA, United States

Bacterial cold water disease (BCWD) is an important disease in rainbow trout aquaculture. Previously, we have identified and validated two major QTL (quantitative trait loci) for BCWD resistance, located on chromosomes Omy08 and Omy25, in the odd-year Troutlodge May spawning population. We also demonstrated that marker-assisted selection (MAS) for BCWD resistance using the favorable haplotypes associated with the two major QTL is feasible. However, each favorable haplotype spans a large genomic region of 1.3–1.6 Mb. Recombination events within the haplotype regions will result in new haplotypes associated with BCWD resistance, which will reduce the accuracy of MAS for BCWD resistance over time. The objectives of this study were 1) to identify additional SNPs (single nucleotide polymorphisms) associated with BCWD resistance using whole-genome sequencing (WGS); 2) to validate the SNPs associated with BCWD resistance using family-based association mapping; 3) to refine the haplotypes associated with BCWD resistance; and 4) to evaluate MAS for BCWD resistance using the refined QTL haplotypes. Four consecutive generations of the Troutlodge May spawning population were evaluated for BCWD resistance. Parents and offspring were sequenced as individuals and in pools based on their BCWD phenotypes. Over 12 million SNPs were identified by mapping the sequences from the individuals and pools to the reference genome. SNPs with significantly different allele frequencies between the two BCWD phenotype groups were selected to develop SNP assays for family-based association mapping in three consecutive generations of the Troutlodge May spawning population. Among the 78 SNPs derived from WGS, 77 SNPs were associated with BCWD resistance in at least one of the three consecutive generations. The additional SNPs associated with BCWD resistance allowed us to reduce the physical sizes of haplotypes associated with BCWD resistance to less than 0.5 Mb. We also demonstrated that the refined QTL haplotypes can be used for MAS in the Troutlodge May spawning population. Therefore, the SNPs and haplotypes reported in this study provide additional resources for improvement of BCWD resistance in rainbow trout.

**Keywords: rainbow trout, bacterial cold water disease, haplotype, SNP, MAS, QTL, whole genome sequencing (WGS)**

OPEN ACCESS

***Edited by:***
*Nguyen Hong Nguyen, University of the Sunshine Coast, Australia*

***Reviewed by:***
*Antti Kause, Natural Resources Institute Finland (Luke), Finland*
*Yulin Jin, Emory University, United States*

****Correspondence:***
*Sixin Liu*
*sixin.liu@usda.gov*

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

***Received:*** *05 May 2022*
***Accepted:*** *06 June 2022*
***Published:*** *23 June 2022*

***Citation:***
*Liu S, Martin KE, Gao G, Long R, Evenhuis JP, Leeds TD, Wiens GD and Palti Y (2022) Identification of Haplotypes Associated With Resistance to Bacterial Cold Water Disease in Rainbow Trout Using Whole-Genome Resequencing. Front. Genet. 13:936806. doi: 10.3389/fgene.2022.936806*

## INTRODUCTION

The global demand for seafood has roughly doubled since the start of the 21st century, and will likely double again by 2050 (Naylor et al., 2021). Rainbow trout (*Oncorhynchus mykiss*) is one of the most widely cultured cold freshwater fish, with production on every continent except Antarctica. The global production of rainbow trout was about 917,000 tons in 2019 (FAO, 2021). Outbreaks of infectious disease are one of the major challenges for rainbow trout production and welfare. Bacterial cold water disease (BCWD), caused by *Flavobacterium psychrophilum*, is a frequent disease in rainbow trout (Nematollahi et al., 2003; Starliper, 2011; Loch and Faisal, 2015). Commercial vaccines for BCWD are not available yet. Use of licensed antibiotics for BCWD treatment increases production cost and antibiotic resistant pathogens may emerge.

Use of genetic resistance is an effective approach to control BCWD in rainbow trout. A rainbow trout line with improved resistance to BCWD has been developed by using family-based selection (Leeds et al., 2010; Wiens et al., 2013a; Wiens et al., 2018). Recently, multiple studies have demonstrated that genomic selection (GS) can substantially improve the accuracy of selection for BCWD resistance in rainbow trout. Vallejo et al. (2017a) reported that GS using a 57K SNP (single nucleotide polymorphism) genotyping array (Palti et al., 2015a) can double the accuracy of selection for BCWD resistance in a commercial breeding population. To reduce the cost of genotyping, the accuracy of GS for BCWD resistance was evaluated with low-density SNP panels. The accuracy of GS remained substantially higher than pedigree-based selection when using 70 SNPs associated with QTLs (quantitative trait locus) for BCWD resistance (Vallejo et al., 2018). To reduce the cost of BCWD phenotyping, it has recently been reported that the accuracy of GS for BCWD resistance without model retraining in the subsequent generation remained higher than pedigree-based selection (Vallejo et al., 2021).

To fully exploit the genetic resistance to BCWD, extensive genetic mapping studies were conducted to identify and validate QTLs for BCWD resistance in rainbow trout. Fraslin et al. (2018) used both immersion and intramuscular injection methods to evaluate double haploids derived from a cross between two rainbow trout isogenic lines, and 15 QTLs for BCWD resistance were identified. Also, two QTLs for BCWD resistance were identified after a natural disease outbreak on a French farm (Fraslin et al., 2019). At the USDA National Center for Cool and Cold Water Aquaculture, we initially used full-sib mapping families to identify and validate QTL for BCWD resistance (Wiens et al., 2013b; Vallejo et al., 2014a; Vallejo et al., 2014b; Palti et al., 2015b; Liu et al., 2015). With the advancement of genomic resources available in rainbow trout such as a SNP genotyping array (Palti et al., 2015a) and a reference genome (Pearse et al., 2019), we performed a genome-wide association study (GWAS) to detect QTL for BCWD resistance (Vallejo et al., 2017b) in the 2013 generation of the Troutlodge May spawning population. Three QTL for BCWD resistance with moderate-large effects, located on chromosomes Omy03, Omy08, and Omy25, were identified. In a follow-up study the three QTLs were validated in the 2015 generation of the Troutlodge May spawning population (Liu et al., 2018). In the same study it was shown that SNP haplotypes associated with the two major QTL on chromosomes Omy08 and Omy25 can be used for marker-assisted selection (MAS) for BCWD resistance. However, the two favorable haplotypes for the two major QTL on chromosomes Omy08 and Omy25 span regions of 1.3 and 1.6 Mb, respectively. Recombination events within the haplotype regions may result in new haplotypes associated with BCWD resistance, which will reduce the accuracy of MAS for BCWD resistance over time. Thus, it is important to identify and validate additional SNPs associated with the two major QTL for BCWD resistance with a goal to reduce the physical size of haplotypes associated with BCWD resistance.

Whole-genome sequencing (WGS) is a powerful tool to discover SNPs and to identify causative genes for traits of interest. With the recent rapid reduction in the cost of next generation sequencing, the use of WGS has become more common in genetic studies of salmonids (Gao et al., 2018; Narum et al., 2018; Thompson et al., 2020; Liu et al., 2021). The objectives of this study were 1) to identify additional SNPs associated with BCWD resistance using WGS; 2) to validate the SNPs associated with BCWD resistance using family-based association mapping; 3) to refine the haplotypes associated with BCWD resistance; and 4) to evaluate MAS for BCWD resistance using the refined QTL haplotypes.

## MATERIALS AND METHODS

### Five Consecutive Generations of the Troutlodge May Spawning Strain

Troutlodge, Inc., has four rainbow trout strains (Liu et al., 2017) named by their peak spawning months. All samples used in this study were from the May spawning strain. Five consecutive generations (**Table 1**) were used in this study. The samples used for WGS were from the 2013 and 2015 generations. Selected families from three consecutive generations, 2015, 2017, and 2019, were used for association analyses of BCWD resistance. Fish of the 2021 generation were used to evaluate MAS for BCWD resistance. Previously, each Troutlodge strain had two populations, odd-year and even-year populations. The odd-year and even-year May spawning populations have been merged into one population since the 2019 generation.

### BCWD Challenge Experiments

BCWD challenge experiments were conducted in four consecutive generations, 2015, 2017, 2019, and 2021 (**Supplementary Table S1**). Fish (80–99 days post-hatch) were challenged by intraperitoneal injection of *Flavobacterium psychrophilum* strain CSF259-93 using the established protocol described in detail by Hadidi et al. (2008). Mortalities were collected daily for 21 days after intraperitoneal injection. Both survival days (DAYS), the number of days to death after BCWD challenge, and survival status (STATUS), 2 for dead fish and 1 for survivors at day 21, were recorded for each fish. Each family of the

**TABLE 1 |** Summary of samples from five consecutive generations of the Troutlodge May spawning population used in this study.

| Generation | Study | Comment |
|---|---|---|
| 2013 | WGS | Selected parents for the 2015 generation were used for WGS (Individuals and pools; N = 20) |
| 2015 | WGS, SNP validation | 1) Selected parents (Individuals and pools; N = 20) and families (pools; N = 480) were used for WGS; 2) Validate SNPs associated with BCWD resistance using 60 families (10 offspring per family; N = 600) with intermediate BCWD survival rates |
| 2017 | SNP validation | Validate SNPs associated with BCWD resistance using 60 families (10 offspring per family; N = 600) with intermediate BCWD survival rates |
| 2019 | SNP validation | Validate SNPs associated with BCWD resistance using 9 families (92 offspring per family; N = 828) with intermediate BCWD survival rates |
| 2021 | MAS | Evaluation of MAS using Pooled BCWD challenge (50 families; 20 fish per family; N = 1,000) |

2015 and 2017 generations was evaluated for BCWD resistance using two replicate tanks (3 L tank with a water flow rate of 1 L per minute) with 40 fish per tank, and the details have been reported in our previous publications (Vallejo et al., 2017a; Liu et al., 2018). For the 2019 generation, we increased the number of fish challenged per family to 3 or 4 replicate tanks (40 fish per tank) based on fish availability. The 2021 families were pooled and challenged as described below to evaluate MAS for BCWD resistance.

## Sequencing of 40 Parents of the 2015 and 2017 Generations

Based on the BCWD survival rates of the 2015 and 2017 families and parental haplotypes for the two targeted QTL regions, Omy08 and Omy25, we selected 40 (**Supplementary Table S2**) parents for individual sequencing and pooled sequencing. First, we sorted the families within each generation by BCWD survival rate from high to low. The parents of top 20 families were assigned to a BCWD resistant (R) group, and the parents of bottom 20 families were assigned to a BCWD susceptible (S) group. Then, the QTL haplotypes of these parents were reconstructed for the two QTL regions using the same SNP sets and method reported in Liu et al. (2018). The parents of the 2015 and 2017 generations were selected to target for the Omy08 and Omy25 BCWD QTL, respectively. For each generation, we selected 10 R parents that are fixed for the favorable haplotype for the targeted QTL and have at least one favorable haplotype for the other QTL. We also selected 10 S parents without any favorable haplotype for the targeted QTL. Thus, a total of 40 parents were selected for WGS with a targeted genome coverage of 15x per sample. In addition to sequencing of individuals, we also pooled equal amount of DNA per fish by BCWD groups within each generation for pooled sequencing. The targeted genome coverage per pool was 30x. The raw sequences of the parents were deposited in sequence read archive (SRA) under BioProject PRJNA681179.

## Sequencing of Pooled Offspring of the 2015 Generation

The sequencing of parents described above might be biased because both BCWD survival rates and QTL haplotypes were used to select the samples used for sequencing. Thus, we decided to use BCWD phenotype as the only criteria to select samples for additional pooled sequencing. Among the 138 families of the 2015 generation evaluated for BCWD resistance, we selected 60 families with intermediate BCWD survival rates that ranged from 24% to 71% for sequencing. For each of the 60 families, we selected the first four fish that died after day 3 (to avoid fish died from injection injury or stress) and four random survivors. Each of the four dead fish or survivors was randomly assigned to one of the four corresponding BCWD status pools. In total, we had four DNA pools of dead fish (S pools) and four DNA pools of survivors (R pools). Equal amount of DNA per sample was pooled for sequencing with a targeted genome coverage of 45x per DNA pool. The sequences of the pooled offspring were deposited under BioProject PRJNA830380.

## Whole-Genome Sequencing and SNP Identification

DNA was extracted from fin clips following the manufacturer's recommended protocols for AutoGenprep 965 (Autogen, Holliston, MA, United States). Whole-genome DNA sequencing libraries were prepared using the KAPA HyperPrep kit (KAPA Biosystems, Wilmington MA), and were sequenced in paired-end (2 × 150 bp) mode on an Illumina HiSeq X sequencer. The sequence reads were mapped to rainbow trout reference genome GCF_002163495.1 (Pearse et al., 2019) using BWA-MEM algorithm (Li, 2013), and alignments were converted to BAM (Binary sequence Alignment/Map) format using SAMtools v1.11 (Li et al., 2009). PCR duplicates were marked and removed using Picard v2.18.2 (http://broadinstitute.github.io/picard/). Following our previously published SNP calling and filtering pipeline (Gao et al., 2018), SNPs were called using program freebayes v1.3.1 (Garrison and Marth, 2012), and were annotated using program SnpEff v4.3 (Cingolani et al., 2012).

## Identification of SNPs in the QTL Regions

For the sequence data of parents for the 2015 and 2017 generations, we used a sliding window approach to identify SNPs in the BCWD QTL regions. We used a window size 10,000 bp and a step size 5,000 bp to calculate the fixation index (Fst) value for each window. For individual sequencing, program VCFtools v0.1.16 (Danecek et al., 2011) was used to calculate Fst value for each sliding window. For pooled sequencing, program PoPoolation2 (Kofler et al., 2011) was used to calculate Fst value for each sliding window. Only

windows with at least 15 SNPs were included to identify windows with significantly different allele frequencies (empirical $p < 0.0001$) between the R and S groups. For individual sequencing, program VCFtools was also used to calculate Fst value for each SNP with MAF (minor allele frequency) greater than 0.05.

To identify SNPs with significantly different allele frequencies between the pools of R and S offspring from the 2015 generation, Fisher's exact tests were performed using the program PoPoolation2 with the following settings: -min-coverage 40, --max-coverage 400 and--min-count 10. To correct for multiple tests, SNPs with $p$-values less than $4.05 \times 10^{-9}$ (Bonferroni-correction for 12,338,978 SNPs) were considered as significant.

## SNP Genotyping

Among the SNPs identified by WGS, we selected a set of SNPs in the Omy08 and Omy25 QTL regions for association analyses. The SNPs were selected for assay design because they met one or more of the following criteria: 1) The Fst values between R and S parents were high; 2) The $p$-values of Fisher's test for different allele frequencies between R and S pools of the 2015 generation were low; 3) The SNPs had high or moderate effects based on SNP annotation; 4) SNPs are near the six SNPs used in our previous haplotype analysis (Liu et al., 2018). The sequences of the selected SNPs were submitted to Fluidigm (South San Francisco, CA) for assay design. After a preliminary evaluation of assay quality using a subset of mapping samples of the 2015 generation, we assembled a panel of 96 SNPs (**Supplementary Table S3**) to genotype the mapping samples of three consecutive generations, 2015, 2017, and 2019.

We followed the SNP genotyping protocol described in our previous study (Liu et al., 2016). Briefly, DNA samples were pre-amplified, and the pre-amplified products were diluted and used for genotyping with 96.96 Dynamic Array IFCs (Integrated Fluidic Circuits). The arrays were read using EP1 system, and genotypes were called automatically using Fluidigm SNP genotyping analysis software 4.1 with a confidence threshold of 85. The genotype clusters were examined by eye for each assay, and any wrong calls or no calls were corrected manually. The computer program PedCheck (O'Connell and Weeks, 1998) was used to identify genotypes with Mendelian inheritance errors between parents and offspring. Seven SNPs (**Supplementary Table S3**) were removed from association analysis due to poor genotype clusters or high rates of genotype discrepancies between parents and offspring.

## Family-Based Association Mapping of BCWD Resistance

The program PLINK 1.9 (Chang et al., 2015) was used for family-based association mapping to validate SNPs associated with BCWD resistance ($p < 0.01$). The procedure QFAM was used to analyze the phenotypic data DAYS, and the PERM option was used to correct the family structure. The procedure TDT (transmission disequilibrium test) was used to analyze the binary phenotype STATUS. The association analyses were performed for each of the three consecutive generations, 2015, 2017, and 2019.

## Haplotype Association Analysis of BCWD Resistance

We used three criteria to select three SNPs per QTL region for haplotype association analysis. 1) The selected SNPs are highly associated with BCWD resistance based on single SNP association analysis; 2) The MAF for each selected SNP is greater than 0.2; and 3) The three SNPs for each QTL region span a genomic region less than 0.5 Mb according to rainbow trout reference genome GCF_002163495.1 (Pearse et al., 2019). Based these three criteria, three SNPs, P489, P194, and P355, were selected for the Omy08 QTL, and three SNPs, P420, P430, and P212, were selected for the Omy25 QTL. The same families used for single SNP analysis described above were also used for haplotype association analysis. The haplotypes for each fish were constructed using Beagle 5.1 (Browning and Browning, 2007), and haplotypes with frequency larger than 0.1 were retained for haplotype association analysis. Program PLINK1.9 was used for haplotype association analysis following the same method for family-based association analysis of single SNP as described above.

## Evaluation of MAS for BCWD Resistance in the 2021 Generation

The parents for the 2021 generation were genotyped with 96 SNPs (**Supplementary Table S3**), and haplotypes for the Omy08 and Omy25 QTL regions were constructed using the same SNPs and method described above. The 163 families of the 2021 generation were sorted by the total count of favorable haplotypes from high to low. The top 25 families were assigned to the high haplotype group, and the bottom 25 families were assigned to low haplotype group. We pooled 10 fish per family by haplotype groups, and the 250 fish were challenged with BCWD in a 40 L tank with a water flow rate of 2.5 L per minute. There were two replicate tanks for each haplotype group. So, a total of 500 fish were challenged per haplotype group. To test the BCWD survival difference between the high and low favorable haplotype groups, a log-rank test was performed using the survival package (Therneau, 2021) available in R version 4.1.2 (R Core Team, 2021).

# RESULTS

## Identification of Genomic Regions Associated With BCWD Resistance Using WGS

The 20 parents for the 2015 generation used for sequencing were selected to target the Omy08 QTL for BCWD resistance. For individual sequencing, all windows with significantly different Fst between R and S parents were in a region between 73.2 and 78.2 Mb on chromosome Omy08 (**Figure 1A**). For pooled sequencing, except two windows on chromosome Omy05 and one window on chromosome Omy13, all the other windows with significantly different Fst between R and S pools were in a region

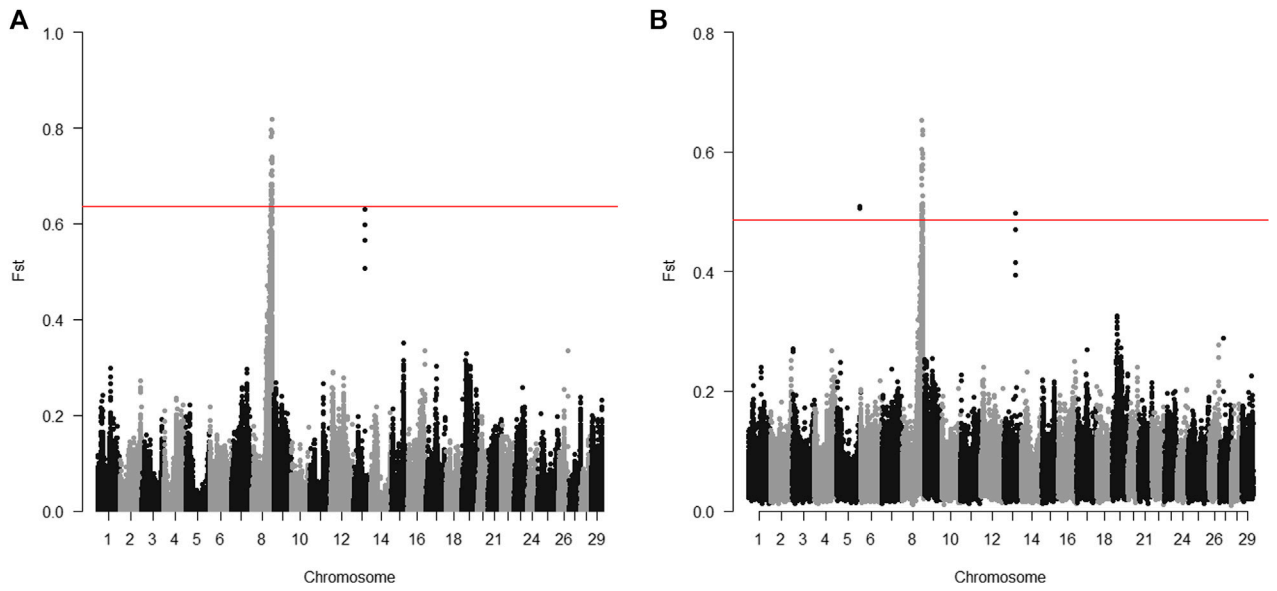



**FIGURE 1 |** Distribution of Fst between R and S groups of parents for the 2015 generation of the Troutlodge May spawning population. **(A)** sequencing of individuals; **(B)** sequencing of pooled samples. The red horizontal lines represent the threshold for significantly different Fst (empirical $p < 0.0001$).

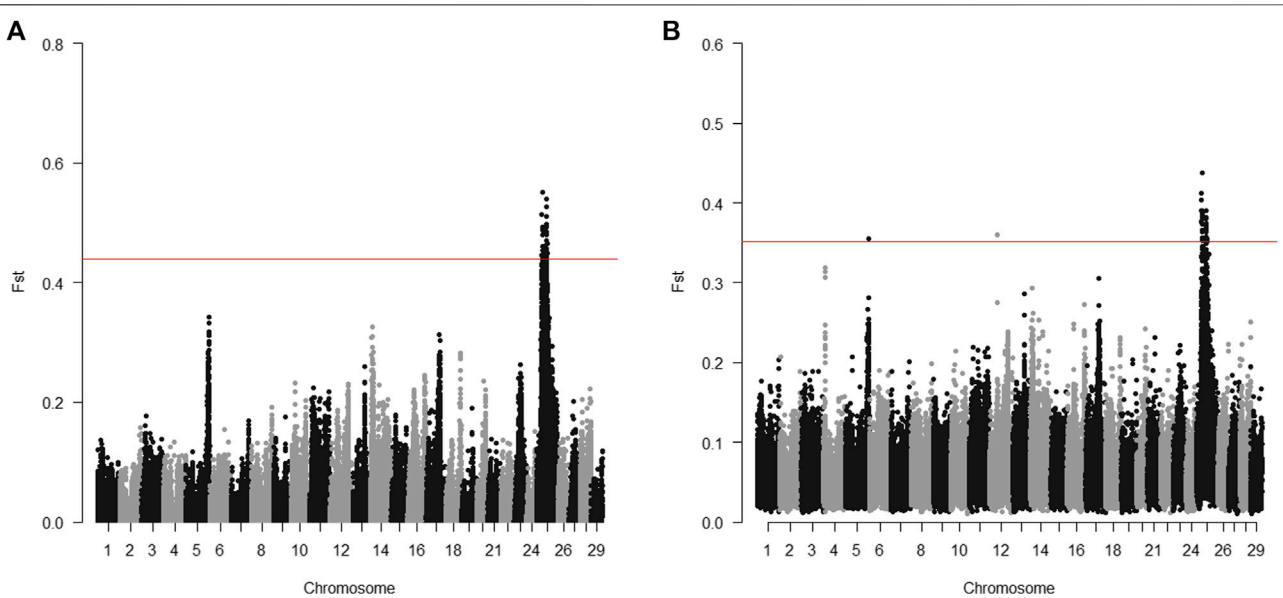



**FIGURE 2 |** Distribution of Fst between R and S groups of parents for the 2017 generation of the Troutlodge May spawning population. **(A)** sequencing of individuals; **(B)** sequencing of pooled samples. The red horizontal lines represent the threshold for significantly different Fst (empirical $p < 0.0001$).

between 73.2 and 78.0 Mb on chromosome Omy08 (**Figure 1B**). Thus, we selected SNPs in the region from 73.2 to 78.2 Mb to validate SNPs associated with the BCWD QTL on chromosome Omy08.

The 20 parents for the 2017 generation used for sequencing were selected to target the Omy25 QTL for BCWD resistance. For individual sequencing, all windows with significantly different Fst between R and S parents were in a region between positions 16.5 and 40.1 Mb on chromosome Omy25 (**Figure 2A**). For pooled sequencing, except one window on chromosome Omy05 and one window on chromosome Omy12, all the other windows with significantly different Fst between R and S pools were located on chromosome Omy25 in a region between 18.9 and 41.0 Mb (**Figure 2B**). Thus, we selected SNPs in the region from 16.5 to 41.0 Mb to validate SNPs associated with the BCWD QTL on chromosome Omy25.

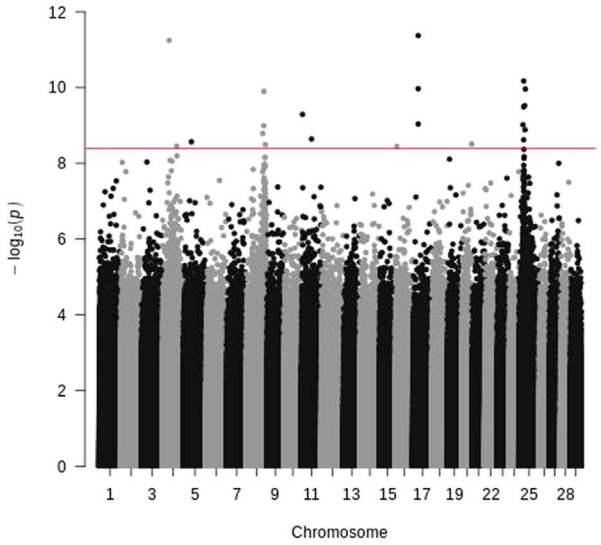



**FIGURE 3 |** Manhattan plot of pooled offspring of the 2015 generation of the Troutlodge May spawning population. The red horizontal line represents the significance threshold (Bonferroni-correction for 12,338,978 SNPs).

To avoid the potential bias of samples used for sequencing just described above, we also sequenced pooled offspring of the 2015 generation. The samples were selected with BCWD phenotypes alone, and they were pooled by survival status. After mapping the sequence reads to the reference genome, 12.3 million SNPs were identified. Based on Fisher's test for each SNP, 21 SNPs with significantly different allele frequencies between the R and S pools were identified (**Figure 3**), and they were located on chromosomes Omy04, Omy05, Omy08, Omy11, Omy16, Omy17, Omy20, and Omy25. Only chromosomes Omy08 and Omy25 had more than three significant SNPs, and these significant SNPs were located in similar QTL regions to those identified by WGS of parents as described above.

## Validation of SNPs Associated With BCWD Resistance

Among the 89 SNPs used for association analysis in three consecutive generations, 2015, 2017, and 2019, 85 SNPs were associated with BCWD resistance in at least one of the three generations (**Table 2**). Also, 77 out of the 85 validated SNPs were identified via WGS reported in this study. Among the 4 SNPs, P161, P176, P316, and P490, that were not associated with BCWD resistance in this study, P490 was the only SNP derived from WGS.

## Refined Haplotypes Associated With BCWD Resistance

The 85 SNPs associated with BCWD resistance validated above allowed us to reduce the physical size of the haplotypes associated with BCWD resistance. Using the three criteria described in the method section, we selected SNPs, P489, P194, and P355, for the Omy08 QTL, and SNPs, P420, P430, and P212, for the Omy25 QTL. The three selected SNPs span a region less than 0.5 Mb. Thus, we reduced the sizes of haplotypes by about two-thirds. The results of haplotype association analysis are summarized in **Table 3**. For the Omy08 QTL, haplotype CGG was associated with BCWD resistance in all three generations, which increases the number of survival days and reduces the risk of death from BCWD. Similarly, haplotype GGG on chromosome Omy25 was also associated with BCWD resistance in all three generations, which increases the number of survival days and reduces the risk of death from BCWD. The other two haplotypes (**Table 3**) were associated with BCWD susceptibility. Because the goal of selective breeding is to improve BCWD resistance, we will focus on the two haplotypes associated with BCWD resistance and refer to them as favorable haplotypes for BCWD resistance.

## Evaluation of MAS for BCWD Resistance

The 25 families from the 2021 generation with high or low counts of favorable haplotypes for the two major QTL regions had an average of 6.48 and 2.56 favorable haplotypes per family, respectively. The BCWD survival rates for the pooled families with high or low counts of favorable haplotypes were 83.6% and 66.2%, respectively. Survival analysis demonstrated that the two BCWD survival curves were significantly different (p = 7e-10) (**Figure 4**) for the two groups of families with high or low counts of favorable haplotypes. Thus, the favorable haplotypes can be used to select families with improved BCWD resistance.

# DISCUSSION

In this study, we used WGS to identify additional SNPs associated with the two major QTL for BCWD resistance, and 77 SNPs identified from WGS were validated by association mapping in three consecutive generations of the Troutlodge May spawning population. The additional SNPs associated with BCWD resistance allowed us to refine the favorable haplotypes associated with BCWD resistance. We demonstrated that the refined favorable QTL haplotypes can be used for MAS for BCWD resistance in the Troutlodge May spawning population.

## Identification of SNPs Associated With BCWD Resistance Using WGS

Among the 78 SNPs derived from WGS, only SNP P490 was not associated with BCWD resistance in this study. The high SNP validation rate was largely due to several factors. 1) The samples used for sequencing were selected with two methods, BCWD phenotypes together with QTL haplotypes or BCWD phenotypes alone; 2) We used two sequencing strategies, sequencing of individuals and pooled samples; 3) The SNPs selected for assay design were filtered by multiple criteria as described in the method section; 4) We removed SNPs with poor genotype quality or SNPs were not associated with BCWD resistance based on a preliminary study with a sub-set of mapping samples. Due to the genotyping platform used in this study, only 96 SNPs including both SNPs derived from WGS and SNPs reported in our previous study (Liu et al., 2018) were used for association mapping. However, analysis of WGS revealed many more SNPs

**TABLE 2 |** Validation of SNPs associated with BCWD resistance in three consecutive generations of the Troutlodge May spawning population.

| Generation | | | | | | 2015 | | | | 2017 | | | | 2019 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Chr.** | **SNP[a]** | **Source** | **Position[b]** | **Allele1** | **Allele2** | **Effect[c]** | ***p* (DAYS)** | **OR[d]** | ***p* (STATUS)** | **Effect** | ***p* (DAYS)** | **OR** | ***p* (STATUS)** | **Effect** | **P (DAYS)** | **OR** | **P (STATUS)** |
| 3 | P178 | Liu et al. (2018) | 57059195 | G | A | 2.8 | 3.7E-04 | 0.6 | 1.1E-04 | 1.3 | NS[e] | 0.7 | 7.1E-03 | 1.9 | NS | 0.8 | NS |
| 8 | P318 | WGS | 75,054,110 | T | G | 3.6 | 7.9E-05 | 0.7 | 1.7E-03 | 1.9 | 2.1E-03 | 0.7 | 2.7E-03 | −0.3 | NS | 1.0 | NS |
| 8 | P319 | WGS | 75,056,552 | G | C | 3.5 | 3.2E-05 | 0.7 | 3.3E-03 | 1.9 | 4.9E-03 | 0.7 | 6.1E-03 | −0.9 | NS | 1.0 | NS |
| 8 | P323 | WGS | 75,064,254 | A | T | 2.8 | 1.6E-03 | 0.7 | NS | 1.9 | 2.3E-03 | 0.7 | 8.5E-03 | −0.3 | NS | 1.0 | NS |
| 8 | **P489** | WGS | 76,457,672 | A | C | −4.0 | 1.0E-04 | 1.4 | NS | -2.9 | 6.3E-05 | 1.4 | NS | −2.2 | NS | 1.3 | 4.5E-03 |
| 8 | P329 | WGS | 76,691,975 | T | A | −4.0 | 3.0E-06 | 1.5 | 2.7E-04 | -2.3 | 1.7E-03 | 1.3 | NS | −2.5 | NS | 1.6 | 1.6E-05 |
| 8 | P330 | WGS | 76,692,073 | T | A | −4.0 | 3.0E-06 | 1.5 | 2.2E-04 | -2.2 | 2.5E-03 | 1.3 | NS | −2.6 | NS | 1.6 | 1.1E-05 |
| 8 | P332 | WGS | 76,696,519 | C | A | −3.9 | 2.0E-06 | 1.4 | 4.8E-03 | -2.9 | 2.4E-05 | 1.4 | NS | −2.1 | NS | 1.2 | NS |
| 8 | **P194** | Liu et al. (2018) | 76,747,151 | A | G | −4.0 | 2.0E-06 | 1.6 | 3.5E-05 | -2.3 | 2.0E-03 | 1.3 | NS | −2.2 | NS | 1.5 | 1.3E-04 |
| 8 | P341 | WGS | 76,758,777 | A | G | −4.0 | 2.0E-06 | 1.5 | 4.8E-04 | −2.6 | 2.4E-04 | 1.4 | 5.0E-03 | −1.2 | NS | 1.2 | NS |
| 8 | P342 | WGS | 76,765,345 | A | T | −3.9 | 3.0E-06 | 1.5 | 5.8E-04 | −2.5 | 6.4E-04 | 1.4 | 7.3E-03 | −2.5 | 8.6E-03 | 1.3 | 5.0E-03 |
| 8 | P346 | WGS | 76,815,705 | G | T | −3.9 | 3.0E-06 | 1.5 | 5.8E-04 | −2.1 | 8.1E-04 | 0.9 | NS | −2.5 | 8.6E-03 | 1.5 | 4.1E-05 |
| 8 | P347 | WGS | 76,844,443 | G | A | −3.8 | 4.2E-05 | 1.4 | 6.8E-03 | −2.4 | 1.2E-03 | 1.3 | NS | −1.8 | NS | 1.3 | NS |
| 8 | P350 | WGS | 76,853,682 | G | T | −4.1 | 1.2E-05 | 1.6 | 3.2E-04 | −2.1 | 5.9E-03 | 1.2 | NS | −1.6 | NS | 1.6 | 3.0E-05 |
| 8 | P354 | WGS | 76,874,914 | C | T | −3.8 | 4.2E-05 | 1.4 | 6.8E-03 | −2.4 | 1.3E-03 | 1.3 | NS | −2.2 | NS | 1.5 | 5.2E-05 |
| 8 | **P355** | WGS | 76,875,468 | A | G | −3.8 | 4.2E-05 | 1.4 | 6.8E-03 | −2.3 | 1.3E-03 | 1.5 | 4.7E-04 | −2.2 | NS | 1.5 | 5.2E-05 |
| 8 | P356 | WGS | 76,879,188 | A | G | −3.8 | 4.2E-05 | 1.4 | 6.8E-03 | −2.4 | 1.3E-03 | 1.3 | NS | −2.2 | NS | 1.5 | 5.2E-05 |
| 8 | P357 | WGS | 76,879,806 | C | A | −3.8 | 4.2E-05 | 1.4 | 6.8E-03 | −2.4 | 1.3E-03 | 1.3 | NS | −2.2 | NS | 1.5 | 5.4E-05 |
| 8 | P191 | Liu et al. (2018) | 78,064,599 | A | G | −3.1 | 3.2E-04 | 1.2 | NS | −2.8 | 2.3E-03 | 1.3 | NS | −1.2 | NS | 1.3 | NS |
| 13 | P494 | WGS | 42,160,393 | C | A | -4.1 | 2.0E-04 | 1.4 | 7.5E-03 | −2.5 | 6.4E-04 | 1.4 | 7.3E-03 | −2.4 | NS | 1.4 | 7.1E-03 |
| 25 | P381 | WGS | 17,246,875 | C | A | −6.2 | 2.0E-06 | 1.6 | NS | −2.5 | 9.1E-03 | 1.6 | 2.4E-03 | −3.0 | 8.6E-03 | 1.5 | 4.2E-05 |
| 25 | P382 | WGS | 17,272,388 | A | G | −5.7 | 3.0E-06 | 1.5 | NS | −2.6 | 1.7E-03 | 1.6 | 7.5E-04 | −2.8 | 8.6E-03 | 1.5 | 2.4E-05 |
| 25 | P383 | WGS | 17,893,051 | T | C | 2.5 | NS | 0.7 | NS | −0.8 | NS | 1.2 | NS | 1.4 | NS | 0.8 | 8.3E-03 |
| 25 | P384 | WGS | 17,944,754 | T | A | 3.7 | 4.2E-05 | 0.6 | 5.2E-03 | 2.5 | 5.4E-03 | 0.6 | 7.2E-03 | 4.5 | NS | 0.3 | 1.2E-06 |
| 25 | P386 | WGS | 18,159,829 | A | G | 3.7 | 4.2E-05 | 0.6 | 4.1E-03 | 2.1 | 9.5E-03 | 0.7 | NS | 4.5 | NS | 0.3 | 1.2E-06 |
| 25 | P230 | Liu et al. (2018) | 18,192,885 | G | T | 2.5 | NS | 0.7 | NS | 2.6 | 5.1E-04 | 0.6 | 2.1E-04 | 2.7 | NS | 0.6 | 2.7E-06 |
| 25 | P391 | WGS | 19,063,776 | A | G | −5.5 | 3.0E-06 | 2.1 | 8.0E-06 | −3.0 | 1.4E-04 | 1.7 | 3.1E-04 | −3.0 | NS | 1.6 | 1.7E-06 |
| 25 | P392 | WGS | 19,131,282 | A | G | −6.3 | 1.0E-06 | 2.0 | 5.4E-08 | −3.7 | 3.0E-06 | 1.7 | 3.5E-05 | −2.6 | NS | 1.5 | 1.3E-06 |
| 25 | P393 | WGS | 19,285,578 | T | C | −6.7 | 1.0E-06 | 1.9 | 1.7E-05 | −3.4 | 1.5E-04 | 1.6 | 1.2E-03 | −2.6 | NS | 1.4 | 2.1E-04 |
| 25 | P395 | WGS | 19,356,106 | A | T | −5.5 | 1.0E-06 | 1.9 | 2.8E-05 | −2.9 | 1.9E-04 | 1.6 | 3.1E-04 | −3.4 | 2.3E-03 | 1.7 | 2.8E-09 |
| 25 | P398 | WGS | 19,480,661 | A | G | 3.9 | 3.0E-05 | 0.6 | 1.4E-03 | 2.8 | 1.5E-03 | 0.6 | 7.2E-03 | 4.8 | NS | 0.3 | 4.2E-07 |
| 25 | P399 | WGS | 19,525,929 | G | T | -6.3 | 1.0E-06 | 2.1 | 5.9E-08 | −3.8 | 1.0E-06 | 1.7 | 2.0E-05 | −2.7 | NS | 1.5 | 1.1E-06 |
| 25 | P214 | Liu et al. (2018) | 19,553,268 | C | A | 4.1 | 1.0E-06 | 0.7 | 1.1E-03 | 2.8 | 2.0E-04 | 0.7 | 6.9E-03 | 3.2 | 3.8E-03 | 0.6 | 1.4E-08 |
| 25 | P402 | WGS | 19,652,654 | T | A | -6.1 | 1.0E-06 | 2.0 | 2.1E-07 | -3.6 | 1.0E-06 | 1.7 | 2.6E-05 | -3.3 | 2.3E-03 | 1.7 | 4.0E-09 |
| 25 | P404 | WGS | 19,682,358 | G | A | 4.1 | 1.0E-06 | 0.7 | 1.1E-03 | 2.9 | 8.1E-05 | 0.7 | 5.0E-03 | 3.5 | 7.2E-03 | 0.6 | 4.8E-09 |
| 25 | P409 | WGS | 19,886,570 | C | T | 3.9 | 1.4E-05 | 0.6 | 2.0E-04 | 1.2 | NS | 0.8 | NS | 1.5 | NS | 0.7 | NS |
| 25 | P412 | WGS | 20059482 | C | A | −5.9 | 1.0E-06 | 1.9 | 1.1E-06 | −3.6 | 4.0E-06 | 1.7 | 1.2E-04 | −2.6 | NS | 1.5 | 5.1E-07 |
| 25 | P413 | WGS | 20,082,809 | T | C | −5.5 | 1.0E-06 | 1.9 | 5.7E-05 | −3.0 | 1.9E-04 | 1.6 | 3.1E-04 | −3.6 | 4.3E-03 | 1.8 | 5.9E-09 |
| 25 | P415 | WGS | 20,153,489 | C | T | 4.1 | 1.5E-05 | 0.6 | 1.6E-03 | 1.6 | NS | 0.7 | 7.2E-03 | 3.3 | NS | 0.5 | 2.1E-05 |
| 25 | P417 | WGS | 20,228,457 | C | T | 3.5 | 7.0E-06 | 0.7 | 4.1E-03 | 1.8 | NS | 0.9 | NS | 2.4 | NS | 0.7 | 8.6E-05 |
| 25 | **P420** | WGS | 20,306,455 | G | A | 4.0 | 1.0E-06 | 0.7 | 5.8E-03 | 3.2 | 2.1E-05 | 0.8 | NS | 3.6 | 7.2E-03 | 0.5 | 1.0E-09 |
| 25 | P495 | WGS | 20,307,827 | C | A | −6.0 | 1.1E-05 | 1.9 | 4.4E-05 | −3.7 | 1.0E-06 | 1.7 | 1.2E-04 | −2.5 | NS | 1.5 | 2.2E-06 |
| 25 | P424 | WGS | 20,403,696 | C | T | −5.9 | 1.0E-06 | 1.9 | 1.1E-06 | −3.6 | 1.0E-06 | 1.7 | 1.2E-04 | −2.5 | NS | 1.5 | 4.9E-07 |
| 25 | P425 | WGS | 20,437,940 | C | T | −6.2 | 2.9E-05 | 1.6 | NS | −2.8 | 3.0E-03 | 1.6 | 4.6E-03 | −3.2 | NS | 1.6 | 8.9E-05 |
| 25 | P426 | WGS | 20,465,120 | G | A | −5.7 | 2.9E-05 | 1.6 | NS | −2.6 | NS | 1.4 | NS | −2.5 | NS | 0.9 | NS |
| 25 | P429 | WGS | 20,570,095 | C | T | −6.6 | 8.5E-06 | 1.8 | 4.3E-03 | −3.4 | NS | 1.3 | NS | 0.0 | NS | 1.1 | NS |

(Continued on following page)

**TABLE 2 |** (*Continued*) Validation of SNPs associated with BCWD resistance in three consecutive generations of the Troutlodge May spawning population.

| Generation | | | | | | 2015 | | | | 2017 | | | | 2019 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chr. | SNP[a] | Source | Position[b] | Allele1 | Allele2 | Effect[c] | *p* (DAYS) | OR[d] | *p* (STATUS) | Effect | *p* (DAYS) | OR | *p* (STATUS) | Effect | P (DAYS) | OR | P (STATUS) |
| 25 | **P430** | WGS | 20,591,515 | T | G | –5.1 | 9.4E-05 | 1.9 | 1.6E-04 | –2.9 | 2.3E-04 | 1.6 | 1.4E-03 | –3.3 | NS | 1.8 | 1.2E-07 |
| 25 | P431 | WGS | 20,647,628 | A | G | –5.1 | 9.4E-05 | 1.9 | 1.6E-04 | –2.9 | 2.3E-04 | 1.6 | 1.4E-03 | –3.4 | NS | 1.8 | 4.6E-08 |
| 25 | P433 | WGS | 20,722,720 | A | G | –5.9 | 1.0E-06 | 1.9 | 1.1E-06 | –3.6 | 1.0E-06 | 1.7 | 1.2E-04 | –2.6 | NS | 1.5 | 1.5E-06 |
| 25 | **P212** | Liu et al. (2018) | 20,751,780 | T | G | –5.9 | 1.0E-06 | 1.9 | 1.6E-06 | –3.6 | 1.0E-06 | 1.6 | 1.5E-04 | –2.6 | NS | 1.5 | 1.5E-06 |
| 25 | P435 | WGS | 20,770,014 | T | G | –6.0 | 1.0E-06 | 1.9 | 1.1E-06 | –3.7 | 1.0E-06 | 1.7 | 1.2E-04 | –2.6 | NS | 1.5 | 1.4E-06 |
| 25 | P436 | WGS | 20,797,381 | C | A | –4.0 | 1.3E-04 | 1.6 | 1.1E-04 | –3.1 | 7.0E-06 | 1.4 | 9.2E-04 | –2.1 | NS | 1.4 | 2.2E-05 |
| 25 | P437 | WGS | 20,814,089 | G | A | –5.6 | 1.0E-06 | 1.8 | 1.3E-04 | –3.1 | 8.1E-05 | 1.7 | 1.8E-04 | –3.5 | 4.3E-03 | 1.7 | 3.3E-09 |
| 25 | P438 | WGS | 20,905,605 | T | C | –6.1 | 1.0E-06 | 1.8 | 3.1E-06 | –3.8 | 1.0E-06 | 1.8 | 3.2E-07 | –3.2 | NS | 1.7 | 1.5E-09 |
| 25 | P439 | WGS | 20,918,179 | C | T | –5.1 | 9.4E-05 | 1.9 | 1.6E-04 | –2.9 | 2.0E-04 | 1.6 | 1.4E-03 | –3.2 | NS | 1.7 | 4.4E-08 |
| 25 | P440 | WGS | 20,937,844 | A | T | 3.9 | 3.0E-05 | 0.6 | 2.4E-03 | 2.5 | 5.9E-03 | 0.7 | NS | 4.8 | NS | 0.3 | 4.2E-07 |
| 25 | P443 | WGS | 21,034,905 | T | G | 4.2 | 1.2E-05 | 0.6 | 2.1E-03 | 1.4 | NS | 0.7 | 7.2E-03 | 4.9 | NS | 0.3 | 1.4E-07 |
| 25 | P446 | WGS | 21,118,315 | A | C | –5.5 | 3.0E-06 | 1.8 | 1.3E-04 | –2.7 | 5.6E-04 | 1.5 | 6.0E-03 | –3.5 | 4.3E-03 | 1.7 | 1.2E-08 |
| 25 | P228 | Liu et al. (2018) | 21,146,360 | T | G | –5.6 | 2.0E-06 | 1.9 | 8.7E-05 | –3.0 | 8.7E-05 | 1.7 | 1.8E-04 | –3.4 | 2.3E-03 | 1.7 | 2.1E-09 |
| 25 | P447 | WGS | 21,164,811 | A | G | –6.2 | 2.9E-05 | 1.6 | NS | –2.8 | 4.1E-03 | 1.6 | 5.8E-03 | –3.2 | NS | 1.6 | 1.2E-04 |
| 25 | P497 | WGS | 21,245,059 | T | C | 3.0 | 2.9E-03 | 0.7 | NS | 2.7 | 3.1E-03 | 0.7 | NS | 4.9 | NS | 0.3 | 1.4E-07 |
| 25 | P448 | WGS | 21,245,244 | T | C | 3.8 | 7.4E-05 | 0.6 | 2.4E-03 | 2.7 | 3.1E-03 | 0.7 | NS | 2.5 | NS | 0.5 | 6.0E-04 |
| 25 | P498 | WGS | 21,260,524 | T | G | 3.6 | 3.5E-04 | 0.8 | NS | 2.9 | 2.0E-05 | 0.7 | 1.3E-04 | 3.4 | 4.0E-03 | 0.6 | 9.4E-09 |
| 25 | P499 | WGS | 21,260,707 | C | T | 3.0 | 3.2E-03 | 0.7 | NS | 2.6 | 5.0E-03 | 0.7 | NS | 4.9 | NS | 0.3 | 9.0E-08 |
| 25 | P451 | WGS | 21,429,395 | T | C | –6.4 | 6.5E-06 | 1.7 | 4.8E-03 | –3.7 | 3.9E-03 | 1.5 | NS | –3.8 | NS | 1.7 | 2.0E-03 |
| 25 | P360 | WGS | 21,430,497 | G | T | 3.4 | 1.5E-05 | 0.7 | 6.2E-03 | 1.5 | NS | 0.9 | NS | 2.7 | NS | 0.7 | 1.0E-05 |
| 25 | P361 | WGS | 21,431,434 | C | T | 3.7 | 2.0E-06 | 0.7 | 9.3E-03 | 3.0 | 1.3E-04 | 0.8 | NS | 3.0 | NS | 0.6 | 2.9E-08 |
| 25 | P363 | WGS | 21,468,711 | C | G | 3.8 | 2.0E-06 | 0.7 | NS | 2.7 | 3.7E-04 | 0.9 | NS | 2.9 | NS | 0.6 | 1.8E-08 |
| 25 | P372 | WGS | 21,469,214 | T | G | 3.9 | 2.0E-06 | 0.7 | 8.3E-03 | 3.4 | 1.4E-05 | 0.7 | NS | 3.5 | NS | 0.6 | 2.7E-10 |
| 25 | P222 | Liu et al. (2018) | 21,530,601 | A | C | 2.5 | NS | 0.5 | NS | 2.8 | 3.4E-03 | 0.7 | NS | 4.9 | NS | 0.3 | 9.0E-08 |
| 25 | P453 | WGS | 21,535,437 | A | G | –6.2 | 2.9E-05 | 1.6 | NS | –2.9 | 2.9E-03 | 1.6 | 4.6E-03 | –3.2 | NS | 1.6 | 6.0E-05 |
| 25 | P458 | WGS | 21,955,099 | T | A | 3.7 | 1.3E-04 | 0.6 | 2.4E-03 | 2.5 | 2.2E-03 | 0.7 | NS | 4.9 | NS | 0.3 | 4.2E-07 |
| 25 | P374 | WGS | 22,346,077 | G | T | 4.0 | 2.0E-06 | 0.7 | 6.0E-03 | 3.3 | 9.0E-06 | 0.8 | NS | 3.6 | 7.2E-03 | 0.6 | 4.4E-09 |
| 25 | P369 | WGS | 22,363,310 | T | G | 3.6 | 3.0E-06 | 0.7 | 9.5E-03 | 3.3 | 1.9E-05 | 0.8 | NS | 3.2 | NS | 0.6 | 4.4E-08 |
| 25 | P371 | WGS | 22,373,578 | A | C | 3.6 | 6.0E-06 | 0.7 | 6.1E-03 | 2.2 | 4.0E-03 | 0.8 | NS | 3.3 | NS | 0.6 | 5.9E-07 |
| 25 | P468 | WGS | 22,900,574 | A | C | 3.3 | 2.7E-05 | 0.7 | NS | 2.3 | 1.5E-03 | 0.9 | NS | 3.5 | 7.2E-03 | 0.6 | 1.1E-07 |
| 25 | P469 | WGS | 22,929,112 | G | A | –2.8 | 8.7E-04 | 1.4 | 4.1E-03 | –1.7 | NS | 1.2 | NS | -3.6 | NS | 1.6 | 1.9E-05 |
| 25 | P470 | WGS | 22,954,748 | T | G | –5.4 | 2.1E-05 | 1.5 | NS | –3.4 | 6.3E-05 | 1.7 | 1.4E-04 | -3.4 | 4.0E-03 | 1.6 | 3.4E-07 |
| 25 | P472 | WGS | 23,514,396 | T | C | 3.4 | 5.6E-05 | 0.6 | 1.8E-04 | 0.8 | NS | 0.8 | NS | -1.0 | NS | 1.1 | NS |
| 25 | P473 | WGS | 23,533,980 | T | C | 4.3 | 1.0E-06 | 0.6 | 1.1E-05 | 2.3 | 1.8E-03 | 0.6 | 6.1E-04 | 2.9 | NS | 0.5 | 1.7E-05 |
| 25 | P476 | WGS | 23,741,596 | C | T | 3.8 | 6.2E-05 | 0.6 | 2.4E-03 | 2.6 | 1.1E-03 | 0.7 | NS | 4.5 | NS | 0.3 | 2.4E-06 |
| 25 | P481 | WGS | 24,385,764 | A | G | 3.9 | 2.9E-05 | 0.6 | 2.4E-03 | 2.7 | 1.1E-03 | 0.7 | NS | 4.9 | NS | 0.3 | 1.8E-06 |
| 25 | P482 | WGS | 24,459,019 | A | C | 4.4 | 1.0E-06 | 0.6 | 1.1E-05 | 2.3 | 3.0E-03 | 0.7 | 1.2E-03 | 2.9 | NS | 0.5 | 3.3E-05 |
| 25 | P483 | WGS | 25,069,717 | A | T | –4.9 | 6.7E-05 | 1.8 | 9.4E-04 | –2.0 | NS | 1.3 | NS | -2.3 | NS | 1.5 | NS |
| 25 | P485 | WGS | 25,580,318 | T | G | 3.9 | 1.3E-05 | 0.6 | 1.6E-03 | 2.0 | NS | 0.8 | NS | 2.5 | NS | 0.6 | 8.8E-03 |

[a]*SNPs, used for haplotype analysis (**Table 3**) are highlighted in bold.*

[b]*SNP, position on rainbow trout reference genome GCF_002163495.1.*

[c]*Allele substitution effect, positive number indicates that allele 1 increases the number of survival days, and negative number indicates that allele 1 reduces the number of survival days.*

[d]*Odds ratio, greater than one indicates that allele 1 increases the risk of death from BCWD, and less than one indicates that allele 1 reduces the risk of death from BCWD.*

[e]*Not significant (p > 0.01).*

**TABLE 3 |** Haplotype association analysis of BCWD resistance in three consecutive generations of the Troutlodge May spawning population.

| Generation | | 2015 | | | | 2017 | | | | 2019 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chr. | Haplotype | Effect[a] | *p* (DAYS) | OR[b] | *p* (STATUS) | Effect | P (DAYS) | OR | P (STATUS) | Effect | P (DAYS) | OR | P (STATUS) |
| 8 | CGG | 3.8 | 4.0E-06 | 0.6 | 2.9E-04 | 2.1 | 2.9E-03 | 0.7 | 1.5E-02 | 2.8 | 7.6E-03 | 0.7 | 9.9E-04 |
| 8 | AAA | −3.9 | 2.7E-05 | 1.2 | NS[c] | −2.8 | 2.3E-03 | 1.1 | NS | −1.3 | NS | 1.0 | NS |
| 25 | GGG | 4.0 | 1.0E-06 | 0.7 | 4.1E-03 | 3.0 | 4.6E-05 | 0.8 | 1.4E-02 | 3.5 | 8.8E-03 | 0.6 | 3.1E-09 |
| 25 | ATT | −5.1 | 7.3E-05 | 1.9 | 1.6E-04 | −2.9 | 1.6E-04 | 1.6 | 1.9E-03 | −3.2 | NS | 1.7 | 1.1E-06 |

[a]*Positive number indicates that the haplotype increases the number of survival days, and negative number indicates that the haplotype reduces the number of survival days.*
[b]*Odds ratio, greater than one indicates that the haplotype increases the risk of death from BCWD, and less than one indicates that the haplotype reduces the risk of death from BCWD.*
[c]*Not significant (p > 0.05).*

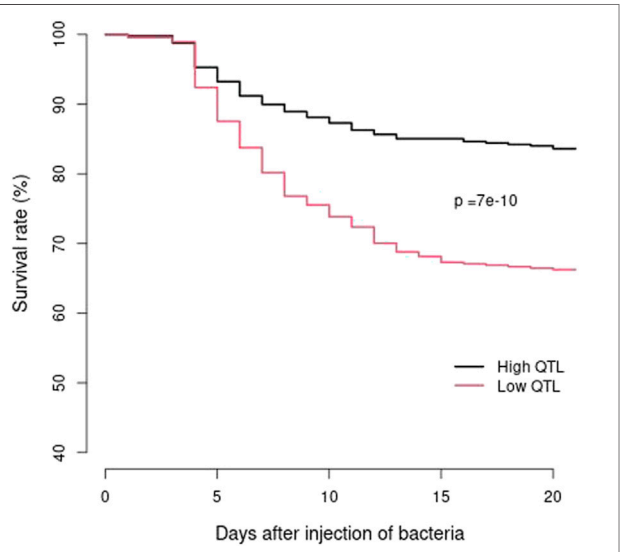



**FIGURE 4 |** BCWD survival curves of the 2021 generation of the Troutlodge May spawning population. The black curve represents the families with high counts of favorable haplotypes for the two major BCWD QTL regions. The red curve represents the families with low counts of favorable haplotypes for the two major BCWD QTL regions.

that were putatively associated with BCWD resistance. Thus, WGS is a powerful tool to identify SNPs associated with BCWD resistance in rainbow trout.

Sequencing pools of individuals (pool-seq) is cost-effective, and has been successfully applied to a variety of studies (Schlotterer et al., 2014). However, pool-seq also has technical challenges and limitations. Unequal representation of DNA samples in the pools can cause false positive signals. The same parental DNA samples were sequenced individually and by pool-seq in this study. Although the genomic regions with significantly different Fst were similar for both sequencing strategies (**Figures 1**, **2**) for the targeted QTL regions, a few additional genomic regions also showed significantly different allele frequencies for the pooled samples, which was likely due to unequal representation of DNA samples in the pools. Compared to the results of sequencing of parents, more genomic regions showed significantly different allele frequencies between the DNA pools of offspring in the 2015 generation (**Figure 3**). Most of them are likely false positives due to the technical challenges of pool-seq. In addition to the possibility of unequal representation of each offspring in the pool, unreliable BCWD phenotypes could be a major factor since it is not possible to have replicated BCWD challenges of an individual fish. The offspring were selected for pool-seq on basis of BCWD survival status of individual fish. On the other hand, the family BCWD survival rates are based on a large number of offspring, and hence are much more reliable than the BCWD survival status of an individual fish.

## Two Robust QTL for BCWD Resistance in Rainbow Trout

QTL validation is essential for implement of MAS in breeding programs and identification of causative genes. The two major QTL for BCWD resistance, located on chromosomes Omy08 and Omy25, were initially identified in the 2013 generation of Troutlodge May spawning population (Vallejo et al., 2017b), and were validated in the 2015 generation of the same population (Liu et al., 2018). In this study, we used WGS to identify additional SNPs associated with these two major QTL, and 77 additional SNPs associated with BCWD resistance were validated in three consecutive generations, 2015, 2017 and 2019, of the Troutlodge May spawning population. Thus, the two major QTL for BCWD resistance are robust in the Troutlodge May population, and it is worthwhile to evaluate MAS for BCWD resistance and to identify positional candidate genes underlying the QTL.

## Robust MAS for BCWD Resistance in Rainbow Trout

We reported previously that the accuracies of retrospective MAS for BCWD resistance using favorable haplotypes associated with the two major BCWD QTL were equal or greater than the accuracies of family-based selection in the same generation of odd-year Troutlodge May spawning population (Liu et al., 2018). In this study, we reduced the physical size of the haplotypes by about two-thirds. We then used the refined haplotypes for MAS for BCWD resistance in the 2021 generation of the Troutlodge May spawning population. Based on the QTL haplotypes of the parents, two groups of families with high or low counts of favorable haplotypes, respectively, were selected for pooled BCWD challenge. The two groups of families had significantly different BCWD survival curves. It is notable that the odd-year and even-

year Troutlodge May spawning populations have been combined into one population since the 2019 generation. Together with the results of retrospective MAS reported previously (Liu et al., 2018), we conclude that MAS for BCWD resistance is robust in the Troutlodge May spawning population.

Although we focused on MAS for BCWD in this study, it is important to note that the additional SNPs associated BCWD resistance reported in this study should also be useful to improve the accuracy of GS for BCWD resistance. We reported previously that the accuracy of GS for BCWD resistance using 70 SNPs associated with BCWD resistance was similar to the accuracy of the whole-genome 57K SNP array (Vallejo et al., 2018). Furthermore, it has been documented that functional and causative variants can be used to improve the accuracy of GS (Xiang et al., 2021). Some of the SNPs reported in this study are located within candidate genes for BCWD resistance (see below).

## Candidate Genes of QTL for BCWD Resistance in Rainbow Trout

Our long-term goal is to identify causative genes for BCWD resistance in rainbow trout. Although the refined haplotypes are associated with resistance to BCWD, they may or may not span the QTL regions. Thus, we arbitrarily extended 0.5 Mb on each end of the refined favorable haplotypes associated with BCWD resistance, and then examined protein-coding genes in the corresponding regions of rainbow trout Arlee reference genome (Gao et al., 2021), which has a better genome coverage than the previous Swanson reference genome (Pearse et al., 2019). Based on the NCBI rainbow trout gene annotation release 101, a total of 70 annotated protein-coding genes were identified in the two major QTL regions (**Supplementary Table S4**).

Among the 40 annotated protein-coding genes in the Omy08 QTL region, multiple genes are likely related to immune responses. Both LOC110530755 and LOC110530756 encode NACHT proteins, which are implicated in apoptosis and MHC (major histocompatibility complex) transcription activation (Koonin and Aravind, 2000; Laing et al., 2008), and play important roles in activation of animal innate immune responses to pathogen infection (Jones et al., 2016). Furthermore, NACHT proteins such as Nod like receptors also play an important role in activation of pyroptosis pathway in both mammals and fish (Song et al., 2022). Three other candidate genes, LOC110530758, LOC110530759, and LOC110530764, encode proteins likely belonging to the signaling lymphocytic activation molecule (SLAM) family of receptors, which in mammals are critical elements for both innate and adaptive immune responses (Veillette, 2006; Cannons et al., 2011). Also, these three SLAM genes were modestly upregulated at day 5 post challenge with *Flavobacterium psychrophilum* in the study reported by Marancik et al. (2015). In addition to the putative functions, the results of association mapping (**Table 2**) also indicated that these genes are strong candidates for the Omy08 QTL. All 8 SNPs in the candidate gene regions (**Supplementary Table S4**) are significantly associated with BCWD resistance in three consecutive generations of the Troutlodge May spawning population (**Table 2**). Thus, these immune-related genes are good candidates for the Omy08 QTL for BCWD resistance.

Among the 30 annotated genes in the Omy25 QTL region, gene LOC100136157, which encodes invariant chain INVX, stands out as a promising candidate gene for this QTL. For simplicity and consistency with rainbow trout literatures, we refer this gene as INVX from now on. Rainbow trout INVX was initially cloned and characterized by Fujiki et al. (2003), and is a homolog of mammalian invariant chain genes, which play important roles in antigen presentation (Schroder, 2016). Transcript level of INVX in rainbow trout cell line culture was significantly increased at 96 and 120 h after immune system activation with PMA (phorbol 12-myristate 13-acetate) (Semple et al., 2019). Also, INVX protein level was significantly reduced at 168 h after PMA stimulation (Semple et al., 2019). In addition to the putative function of INVX, there is also another line of evidence supporting INVX as a candidate gene for the Omy25 QTL. SNP P446, located in the intron of gene INVX, was significantly associated with BCWD resistance in three consecutive generations of the Troutlodge May spawning population (**Table 2**). Therefore, we will continue to evaluate this candidate gene using other approaches in the future.

In addition to the candidate genes highlighted above, we would like to caution that other genes could also be candidate genes for the BCWD QTL. Although we focused on protein-coding genes, there are also two annotated non-coding RNA genes in the Omy08 QTL region. We should not completely rule out other genes in the QTL regions until we can identify with high confidence the causative genes underlying the two QTL for BCWD resistance in rainbow trout.

## CONCLUSION

WGS is a powerful tool to identify additional SNPs associated with the two major QTL for BCWD resistance in rainbow trout. The additional SNPs allowed us to reduce the physical size of haplotypes associated with BCWD resistance. We also demonstrated that the refined favorable QTL haplotypes can be used for MAS for BCWD resistance in the Troutlodge May spawning population. Thus, the additional SNPs and refined haplotypes associated with BCWD resistance reported in this study are useful for improvement of BCWD resistance and for eventual identification of genes for BCWD resistance in rainbow trout.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

## ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Care and Use Committee, National Center for Cool and Cold Water Aquaculture, Agriculture Research Service, United States Department of Agriculture.

## AUTHOR CONTRIBUTIONS

SL and YP conceived and planned the study; KM participated in the study planning and provided pedigree records, germplasm for disease challenges and tissue or DNA samples from the parents; YP, JE, TL, GW and SL coordinated, supervised and performed the disease challenge experiments; GG mapped the sequences to reference genome and called SNPs; SL designed genotyping plans and RL performed the genotyping experiments; SL analyzed the data and drafted the manuscript. All authors read and approved the final manuscript.

## FUNDING

This study was supported by Agricultural Research Service CRIS projects 8082-32000-007 and 8082-31000-013.

## ACKNOWLEDGMENTS

The authors would like to thank Kristy Shewbridge, Ryan Lipscomb, Clayton Birkett, Jenea McGowan, Josh Kretzer, Travis Moreland, Keira Osbourn, Vanessa Panaway, and Joe Beach for fish rearing and help with the disease challenge experiments. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture (USDA). USDA is an equal opportunity provider and employer.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.936806/full#supplementary-material

**Supplementary Table S1 |** Detailed information of BCWD challenge experiments of four consecutive generations of the Troutlodge May spawning population.

**Supplementary Table S2 |** The 40 parents of the 2015 and 2017 generations of the Troutlodge May spawning population selected for sequencing.

**Supplementary Table S3 |** The primer sequences of 96 SNPs used for association mapping in three consecutive generations of the Troutlodge May spawning population.

**Supplementary Table S4 |** Protein-coding genes in the two major BCWD QTL regions based on NCBI rainbow trout gene annotation release 101.

## REFERENCES

Browning, S. R., and Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies by Use of Localized Haplotype Clustering. *Am. J. Hum. Genet.* 81, 1084–1097. doi:10.1086/521987

Cannons, J. L., Tangye, S. G., and Schwartzberg, P. L. (2011). SLAM Family Receptors and SAP Adaptors in Immunity. *Annu. Rev. Immunol.* 29 (29), 665–705. doi:10.1146/annurev-immunol-030409-101302

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *Gigascience* 4, 7. doi:10.1186/s13742-015-0047-8

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of *Drosophila M* Strain W1118; Iso-2; Iso-3. *Fly* 6, 80–92. doi:10.4161/fly.19695

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., Depristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

FAO (2021). Fishery and Aquaculture Statistics. Global Aquaculture Production 1950-2019 (FishstatJ). FAO Fisheries Division. Rome. [Online]. Available at: http://www.fao.org/fishery/statistics/software/fishstatj/en (Accessed September 17, 2021).

Fraslin, C., Dechamp, N., Bernard, M., Krieg, F., Hervet, C., Guyomard, R., et al. (2018). Quantitative Trait Loci for Resistance to Flavobacterium Psychrophilum in Rainbow Trout: Effect of the Mode of Infection and Evidence of Epistatic Interactions. *Genet. Sel. Evol.* 50, 60. doi:10.1186/s12711-018-0431-9

Fraslin, C., Brard-Fudulea, S., D'ambrosio, J., Bestin, A., Charles, M., Haffray, P., et al. (2019). Rainbow Trout Resistance to Bacterial Cold Water Disease: Two New Quantitative Trait Loci Identified after a Natural Disease Outbreak on a French Farm. *Anim. Genet.* 50, 293–297. doi:10.1111/age.12777

Fujiki, K., Smith, C. M., Liu, L., Sundick, R. S., and Dixon, B. (2003). Alternate Forms of MHC Class II-Associated Invariant Chain Are Not Produced by Alternative Splicing in Rainbow Trout (*Oncorhynchus Mykiss*) but Are Encoded by Separate Genes. *Dev. Comp. Immunol.* 27, 377–391. doi:10.1016/s0145-305x(02)00119-2

Gao, G., Nome, T., Pearse, D. E., Moen, T., Naish, K. A., Thorgaard, G. H., et al. (2018). A New Single Nucleotide Polymorphism Database for Rainbow Trout Generated through Whole Genome Resequencing. *Front. Genet.* 9, 147. doi:10.3389/fgene.2018.00147

Gao, G. T., Magadan, S., Waldbieser, G. C., Youngblood, R. C., Wheeler, P. A., Scheffler, B. E., et al. (2021). A Long Reads-Based De-Novo Assembly of the Genome of the Arlee Homozygous Line Reveals Chromosomal Rearrangements in Rainbow Trout. *G3-Genes Genomes Genet.* 11 (4), jkab052. doi:10.1093/g3journal/jkab052

Garrison, E., and Marth, G. (2012). Haplotype-based Variant Detection from Short-Read Sequencing. arXiv:1207.3907 [q-bio.GN].

Hadidi, S., Glenney, G. W., Welch, T. J., Silverstein, J. T., and Wiens, G. D. (2008). Spleen Size Predicts Resistance of Rainbow Trout to Flavobacterium Psychrophilum Challenge. *J. Immunol.* 180, 4156–4165. doi:10.4049/jimmunol.180.6.4156

Jones, J. D., Vance, R. E., and Dangl, J. L. (2016). Intracellular Innate Immune Surveillance Devices in Plants and Animals. *Science* 354 (6316), aaf6395. doi:10.1126/science.aaf6395

Kofler, R., Pandey, R. V., and Schlotterer, C. (2011). PoPoolation2: Identifying Differentiation between Populations Using Sequencing of Pooled DNA Samples (Pool-Seq). *Bioinformatics* 27, 3435–3436. doi:10.1093/bioinformatics/btr589

Koonin, E. V., and Aravind, L. (2000). The NACHT Family - A New Group of Predicted NTPases Implicated in Apoptosis and MHC Transcription Activation. *Trends Biochem. Sci.* 25, 223–224. doi:10.1016/s0968-0004(00)01577-2

Laing, K. J., Purcell, M. K., Winton, J. R., and Hansen, J. D. (2008). A Genomic View of the NOD-Like Receptor Family in Teleost Fish: Identification of a Novel NLR Subfamily in Zebrafish. *BMC Evol. Biol.* 8, 42. doi:10.1186/1471-2148-8-42

Leeds, T. D., Silverstein, J. T., Weber, G. M., Vallejo, R. L., Palti, Y., Rexroad, C. E., et al. (2010). Response to Selection for Bacterial Cold Water Disease Resistance in Rainbow Trout. *J. Animal Sci.* 88, 1936–1946. doi:10.2527/jas.2009-2538

Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. arXiv:1303.3997v2.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Liu, S., Vallejo, R. L., Palti, Y., Gao, G., Marancik, D. P., Hernandez, A. G., et al. (2015). Identification of Single Nucleotide Polymorphism Markers Associated with Bacterial Cold Water Disease Resistance and Spleen Size in Rainbow Trout. *Front. Genet.* 6, 298. doi:10.3389/fgene.2015.00298

Liu, S., Gao, G., Layer, R. M., Thorgaard, G. H., Wiens, G. D., Leeds, T. D., et al. (2021). Identification of High-Confidence Structural Variants in Domesticated Rainbow Trout Using Whole-Genome Sequencing. *Front. Genet.* 12, 639355. doi:10.3389/fgene.2021.639355

Liu, S., Palti, Y., Gao, G., and Rexroad, C. E. (2016). Development and Validation of a SNP Panel for Parentage Assignment in Rainbow Trout. *Aquaculture* 452, 178–182. doi:10.1016/j.aquaculture.2015.11.001

Liu, S., Palti, Y., Martin, K. E., Parsons, J. E., and Rexroad, C. E. (2017). Assessment of Genetic Differentiation and Genetic Assignment of Commercial Rainbow Trout Strains Using a SNP Panel. *Aquaculture* 468, 120–125. doi:10.1016/j.aquaculture.2016.10.004

Liu, S., Vallejo, R. L., Evenhuis, J. P., Martin, K. E., Hamilton, A., Gao, G., et al. (2018). Retrospective Evaluation of Marker-Assisted Selection for Resistance to Bacterial Cold Water Disease in Three Generations of a Commercial Rainbow Trout Breeding Population. *Front. Genet.* 9, 286. doi:10.3389/fgene.2018.00286

Loch, T. P., and Faisal, M. (2015). Emerging Flavobacterial Infections in Fish: A Review. *J. Adv. Res.* 6, 283–300. doi:10.1016/j.jare.2014.10.009

Marancik, D., Gao, G., Paneru, B., Ma, H., Hernandez, A. G., Salem, M., et al. (2015). Whole-Body Transcriptome of Selectively Bred, Resistant-, Control-, and Susceptible-Line Rainbow Trout Following Experimental Challenge with Flavobacterium Psychrophilum. *Front. Genet.* 5, 453. doi:10.3389/fgene.2014.00453

Narum, S. R., Di Genova, A., Micheletti, S. J., and Maass, A. (2018). Genomic Variation Underlying Complex Life-History Traits Revealed by Genome Sequencing in Chinook Salmon. *Proc. R. Soc. B-Biol. Sci.* 285 (1883), 20180935. doi:10.1098/rspb.2018.0935

Naylor, R. L., Kishore, A., Sumaila, U. R., Issifu, I., Hunter, B. P., Belton, B., et al. (2021). Blue Food Demand across Geographic and Temporal Scales. *Nat. Commun.* 12, 5413. doi:10.1038/s41467-021-25516-4

Nematollahi, A., Decostere, A., Pasmans, F., and Haesebrouck, F. (2003). Flavobacterium Psychrophilum Infections in Salmonid Fish. *J. Fish. Dis.* 26, 563–574. doi:10.1046/j.1365-2761.2003.00488.x

O'Connell, J. R., and Weeks, D. E. (1998). PedCheck: A Program for Identification of Genotype Incompatibilities in Linkage Analysis. *Am. J. Hum. Genet.* 63, 259–266. doi:10.1086/301904

Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., et al. (2015a). The Development and Characterization of a 57K Single Nucleotide Polymorphism Array for Rainbow Trout. *Mol. Ecol. Resour.* 15, 662–672. doi:10.1111/1755-0998.12337

Palti, Y., Vallejo, R. L., Gao, G., Liu, S., Hernandez, A. G., Rexroad, C. E., et al. (2015b). Detection and Validation of QTL Affecting Bacterial Cold Water Disease Resistance in Rainbow Trout Using Restriction-Site Associated DNA Sequencing. *PLoS One* 10, e0138435. doi:10.1371/journal.pone.0138435

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., et al. (2019). Sex-Dependent Dominance Maintains Migration Supergene in Rainbow Trout. *Nat. Ecol. Evol.* 3, 1731–1742. doi:10.1038/s41559-019-1044-6

R Core Team (2021). *R: A Language and Environment for Statistical Computing [Online]*. Vienna, Austria: R Foundation for Statistical Computing. Available: https://www.R-project.org (Accessed Dec. 17, 2021).

Schlotterer, C., Tobler, R., Kofler, R., and Nolte, V. (2014). Sequencing Pools of Individuals-Mining Genome-Wide Polymorphism Data without Big Funding. *Nat. Rev. Genet.* 15, 749–763. doi:10.1038/nrg3803

Schröder, B. (2016). The Multifaceted Roles of the Invariant Chain CD74 - More Than Just a Chaperone. *Biochim. Biophys. Acta (BBA) - Mol. Cell Res.* 1863, 1269–1281. doi:10.1016/j.bbamcr.2016.03.026

Semple, S. L., Heath, G., Christie, D., Braunstein, M., Kales, S. C., and Dixon, B. (2019). Immune Stimulation of Rainbow Trout Reveals Divergent Regulation of MH Class II-Associated Invariant Chain Isoforms. *Immunogenetics* 71, 407–420. doi:10.1007/s00251-019-01115-y

Song, Z., Zou, J., Wang, M., Chen, Z., and Wang, Q. (2022). A Comparative Review of Pyroptosis in Mammals and Fish. *J. Inflamm. Res.* 15, 2323–2331. doi:10.2147/jir.s361266

Starliper, C. E. (2011). Bacterial Coldwater Disease of Fishes Caused by *Flavobacterium Psychrophilum*. *J. Adv. Res.* 2, 97–108. doi:10.1016/j.jare.2010.04.001

Therneau, T. M. (2021). A Package for Survival Analysis in R. R Package Version 3.2-13. [Online] Available: https://CRAN.R-project.org/package=survival (Accessed Dec. 17, 2021).

Thompson, N. F., Anderson, E. C., Clemento, A. J., Campbell, M. A., Pearse, D. E., Hearsey, J. W., et al. (2020). A Complex Phenotype in Salmon Controlled by a Simple Change in Migratory Timing. *Science* 370, 609–613. doi:10.1126/science.aba9059

Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (2017a). Genomic Selection Models Double the Accuracy of Predicted Breeding Values for Bacterial Cold Water Disease Resistance Compared to a Traditional Pedigree-Based Model in Rainbow Trout Aquaculture. *Genet. Sel. Evol.* 49, 17. doi:10.1186/s12711-017-0293-6

Vallejo, R. L., Liu, S., Gao, G., Fragomeni, B. O., Hernandez, A. G., Leeds, T. D., et al. (2017b). Similar Genetic Architecture with Shared and Unique Quantitative Trait Loci for Bacterial Cold Water Disease Resistance in Two Rainbow Trout Breeding Populations. *Front. Genet.* 8, 156. doi:10.3389/fgene.2017.00156

Vallejo, R. L., Cheng, H., Fragomeni, B. O., Gao, G., Silva, R. M. O., Martin, K. E., et al. (2021). The Accuracy of Genomic Predictions for Bacterial Cold Water Disease Resistance Remains Higher Than the Pedigree-Based Model One Generation after Model Training in a Commercial Rainbow Trout Breeding Population. *Aquaculture* 545, 737164. doi:10.1016/j.aquaculture.2021.737164

Vallejo, R. L., Palti, Y., Liu, S., Evenhuis, J. P., Gao, G., Rexroad, C. E., et al. (2014b). Detection of QTL in Rainbow Trout Affecting Survival When Challenged with Flavobacterium Psychrophilum. *Mar. Biotechnol.* 16, 349–360. doi:10.1007/s10126-013-9553-9

Vallejo, R. L., Palti, Y., Liu, S., Marancik, D. P., and Wiens, G. D. (2014a). Validation of Linked QTL for Bacterial Cold Water Disease Resistance and Spleen Size on Rainbow Trout Chromosome Omy19. *Aquaculture* 432, 139–143. doi:10.1016/j.aquaculture.2014.05.003

Vallejo, R. L., Silva, R. M. O., Evenhuis, J. P., Gao, G., Liu, S., Parsons, J. E., et al. (2018). Accurate Genomic Predictions for BCWD Resistance in Rainbow Trout Are Achieved Using Low-Density SNP Panels: Evidence that Long-range LD is a Major Contributing Factor. *J. Anim. Breed. Genet.* 135, 263–274. doi:10.1111/jbg.12335

Veillette, A. (2006). Immune Regulation by SLAM Family Receptors and SAP-Related Adaptors. *Nat. Rev. Immunol.* 6, 56–66. doi:10.1038/nri1761

Wiens, G. D., Lapatra, S. E., Welch, T. J., Evenhuis, J. P., Rexroad, C. E., and Leeds, T. D. (2013a). On-Farm Performance of Rainbow Trout (*Oncorhynchus Mykiss*) Selectively Bred for Resistance to Bacterial Cold Water Disease: Effect of Rearing Environment on Survival Phenotype. *Aquaculture* 388-391, 128–136. doi:10.1016/j.aquaculture.2013.01.018

Wiens, G. D., Palti, Y., and Leeds, T. D. (2018). Three Generations of Selective Breeding Improved Rainbow Trout (*Oncorhynchus Mykiss*) Disease Resistance against Natural Challenge with Flavobacterium Psychrophilum during Early Life-Stage Rearing. *Aquaculture* 497, 414–421. doi:10.1016/j.aquaculture.2018.07.064

Wiens, G. D., Vallejo, R. L., Leeds, T. D., Palti, Y., Hadidi, S., Liu, S., et al. (2013b). Assessment of Genetic Correlation between Bacterial Cold Water Disease Resistance and Spleen Index in a Domesticated Population of Rainbow Trout: Identification of QTL on Chromosome Omy19. *PLoS One* 8, e75749. doi:10.1371/journal.pone.0075749

Xiang, R. D., Macleod, I. M., Daetwyler, H. D., De Jong, G., O'connor, E., Schrooten, C., et al. (2021). Genome-Wide Fine-Mapping Identifies Pleiotropic and Functional Variants that Predict Many Traits across Global Cattle Populations. *Nat. Commun.* 12, 860. doi:10.1038/s41467-021-21001-0

**Conflict of Interest:** KM was employed by the company Troutlodge Inc.

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

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 15 July 2022
doi: 10.3389/fgene.2022.926638

# Genome-Wide Association and Genomic Prediction of Growth Traits in the European Flat Oyster (*Ostrea edulis*)

*Carolina Peñaloza[1], Agustin Barria[1], Athina Papadopoulou[2], Chantelle Hooper[2], Joanne Preston[3], Matthew Green[2], Luke Helmer[3,4,5], Jacob Kean-Hammerson[4], Jennifer C. Nascimento-Schulze[2,6], Diana Minardi[2], Manu Kumar Gundappa[1], Daniel J. Macqueen[1], John Hamilton[7], Ross D. Houston[8]*† and Tim P. Bean[1]*†*

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom, [2]Centre for Environment, Fisheries and Aquaculture Science (Cefas), Weymouth Laboratory, Weymouth, United Kingdom, [3]Institute of Marine Sciences, University of Portsmouth, Portsmouth, United Kingdom, [4]Blue Marine Foundation, London, United Kingdom, [5]Ocean and Earth Science, University of Southampton, Southampton, United Kingdom, [6]College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom, [7]Lochnell Oysters, Oban, United Kingdom, [8]Benchmark Genetics, Penicuik, United Kingdom

The European flat oyster (*Ostrea edulis*) is a bivalve mollusc that was once widely distributed across Europe and represented an important food resource for humans for centuries. Populations of *O. edulis* experienced a severe decline across their biogeographic range mainly due to overexploitation and disease outbreaks. To restore the economic and ecological benefits of European flat oyster populations, extensive protection and restoration efforts are in place within Europe. In line with the increasing interest in supporting restoration and oyster farming through the breeding of stocks with enhanced performance, the present study aimed to evaluate the potential of genomic selection for improving growth traits in a European flat oyster population obtained from successive mass-spawning events. Four growth-related traits were evaluated: total weight (TW), shell height (SH), shell width (SW) and shell length (SL). The heritability of the growth traits was in the low-moderate range, with estimates of 0.45, 0.37, 0.22, and 0.32 for TW, SH, SW and SL, respectively. A genome-wide association analysis revealed a largely polygenic architecture for the four growth traits, with two distinct QTLs detected on chromosome 4. To investigate whether genomic selection can be implemented in flat oyster breeding at a reduced cost, the utility of low-density SNP panels was assessed. Genomic prediction accuracies using the full density panel were high (> 0.83 for all traits). The evaluation of the effect of reducing the number of markers used to predict genomic breeding values revealed that similar selection accuracies could be achieved for all traits with 2K SNPs as for a full panel containing 4,577 SNPs. Only slight reductions in accuracies were observed at the lowest SNP density tested (i.e., 100 SNPs), likely due to a high relatedness between individuals being included in the training and validation sets during cross-validation. Overall, our results suggest that the genetic improvement of growth traits in oysters is feasible. Nevertheless, and although low-density SNP panels appear as a promising strategy for applying GS at a reduced cost, additional populations with different degrees of genetic relatedness should be assessed to derive estimates of prediction accuracies to be expected in practical breeding programmes.

**Keywords: *Ostrea edulis*, oyster, GWAS, genomic selection, growth, aquaculture**

OPEN ACCESS

***Edited by:***
*Nguyen Hong Nguyen, University of the Sunshine Coast, Australia*

***Reviewed by:***
*Neil Thompson, Agricultural Research Service (USDA), United States*
*Panya Sae-Lim, MOWI Genetics, Norway*

***Correspondence:**
*Tim P. Bean*
*tim.bean@roslin.ed.ac.uk*
*Ross D. Houston*
*ross.houston@bmkgenetics.com*

*†These authors have contributed equally to this work*

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

***Received:*** *22 April 2022*
***Accepted:*** *17 June 2022*
***Published:*** *15 July 2022*

***Citation:***
*Peñaloza C, Barria A, Papadopoulou A, Hooper C, Preston J, Green M, Helmer L, Kean-Hammerson J, Nascimento-Schulze JC, Minardi D, Gundappa MK, Macqueen DJ, Hamilton J, Houston RD and Bean TP (2022) Genome-Wide Association and Genomic Prediction of Growth Traits in the European Flat Oyster (Ostrea edulis). Front. Genet. 13:926638. doi: 10.3389/fgene.2022.926638*

## INTRODUCTION

The European flat oyster (*Ostrea edulis*) was an abundant native bivalve species and an important fishery resource in much of Europe up to the 19th century (Pogoda, 2019). However, populations of *O. edulis* experienced a severe decline across their biogeographic range due to an array of factors including overfishing and habitat degradation (Thurstan et al., 2013), the subsequent invasion of non-native species (e.g., slipper limpet, *Crepidula fornicata*) (Helmer et al., 2019; Preston et al., 2020) and pathogenic diseases (Robert et al., 1991; Sas et al., 2020). The continuous decimation of native populations in the Atlantic and Mediterranean seas led to significant changes in oyster production, which progressively shifted towards farming (Korringa, 1976), and eventually to the cultivation of different species including *Crassostrea angulata* (Boeliguf, 2000) and the non-indigenous Pacific oyster (*Crassostrea gigas*) (Walne and Helm, 1979; Grizel and Héral, 1991). The Pacific oyster was introduced into Europe for aquaculture purposes owing to its favourable production traits, such as a faster growth rate and higher resistance to the main diseases affecting *C. angulata* and *O. edulis* (Grizel and Héral, 1991; Renault et al., 1995). Worldwide oyster production is now dominated by the Pacific oyster (97.7%), while the production of the European flat oyster remains stably low, constituting just ~0.2% of global production in 2002 (FAO, 2022). Despite the demand for shellfish continues to increase (Botta et al., 2020), the level of *O. edulis* production is stagnant. One of the main factors that hinders the growth of the industry is the lack of a substantial and steady supply of oyster seed (i.e., juveniles) (see Colsoul et al. (2021) for a review). Hence, the optimization of oyster larval production in hatcheries and spatting ponds is key for future European flat oyster aquaculture, as well as for restoration projects, which are also expected to rely on sustainable sources of juveniles for restocking (Pogoda et al., 2020). Importantly, the artificial propagation of flat oyster seed will facilitate the application of selective breeding programmes. Although selective breeding programmes are typically used to improve aquaculture production, they could also benefit the ecological restoration of *O. edulis*. If desirable traits such as disease resistance are found to have a strong genetic component, then increased resistance to life-limiting diseases — such as bonamiosis (Naciri-Graven et al., 1998; Culloty et al., 2004) — could potentially be achieved while maintaining the adaptive potential (i.e., genetic diversity) of restored populations.

Selective breeding in oysters has mainly focused on improving meat yield, disease resistance, survival and growth (Toro and Newkirk, 1990; Allen et al., 1993; Ragone Calvo et al., 2003; Ward et al., 2005; Dégremont et al., 2015; De Melo et al., 2016; Proestou et al., 2016; Camara et al., 2017; Zhang et al., 2019), with a recent interest in nutritional content and shell shape (Grizzle et al., 2017; Liu et al., 2019; Meng et al., 2019; Wan et al., 2020; He et al., 2022). Among these traits, growth is comparatively simple to assess and consequently select for using phenotypic information. Although the direct comparison of heritability estimates from different studies is difficult (e.g., due to intrinsic differences between populations), estimates for growth rate in oysters tend to be moderate (e.g., 0.31 and 0.55—Evans and Langdon (2006) and De Melo et al. (2016), respectively). As a result, fast-growing lines of oysters have been developed for some of the main commercial species, such as the Pacific (*C. gigas*) (Zhang et al., 2019), Portuguese (*C. angulata*) (Vu et al., 2020), Eastern (*C. virginica*) (Varney and Wilbur, 2020) and Sydney rock (*Saccostrea glomerata*) (Fitzer et al., 2019) oyster. Initial attempts to genetically improve the European flat oyster *O. edulis* resulted in an average 23% increase in growth rate compared to an unselected (control) line (Newkirk and Haley, 1982). This striking genetic response was not replicated in a second generation of selection, possibly due to unintentional inbreeding (Newkirk and Haley, 1983). Indeed, even relatively modest levels of inbreeding have been shown to significantly affect performance traits in oysters (Evans et al., 2004), highlighting the importance of an adequate management of genetic diversity in hatchery-derived stocks. Moreover, oysters and bivalves in general, appear to have a high genetic load [see for a review Plough (2016)] and, therefore, may be particularly susceptible to inbreeding depression. Hence, the incorporation of genomic tools into shellfish breeding schemes will be key for balancing genetic gain with population diversity in order to sustain the long-term progress for traits under selection.

A vast array of genomic tools and resources have become available for genetic research and breeding applications in oysters. For example, for economically relevant species, chromosome-level genome assemblies (Li et al., 2021; Modak et al., 2021; Peñaloza et al., 2021; Qi et al., 2021), SNP arrays (Lapegue et al., 2014; Gutierrez et al., 2017; Qi et al., 2017) and medium-density linkage maps (Jones et al., 2013; Wang et al., 2016; Gutierrez et al., 2020; Li et al., 2018; Yin et al., 2020) have been produced. These resources have been applied to examine the genetic basis of growth (Jones et al., 2014; Gutierrez et al., 2018; He et al., 2021), low salinity tolerance (McCarty et al., 2022), disease resistance (Gutierrez et al., 2020; Yang et al., 2022) and nutritional content (Meng et al., 2019). For the European flat oyster, high-quality genomes have recently been released (Boutet et al., 2022; Gundappa et al., 2022), which along with available high-throughput genotyping techniques (e.g., SNP arrays and genotype-by-sequencing approaches), provide the opportunity for gaining insight into the genomic architecture of relevant production traits. Most of the traits of economic importance in aquaculture species have a polygenic architecture (Zenger et al., 2019). For polygenic traits (i.e., those controlled by many loci), the application of predictive techniques such as genomic selection (GS) may enable a faster genetic gain than conventional pedigree-based selection. GS is a method based on genome-wide markers in which the effect of all loci are simultaneously used for predicting the estimated breeding values (EBV) of selection candidates (Meuwissen et al., 2001), and has shown major potential in aquaculture species, where it can be used to

characterise variation within and between large families of potential breeders. However, commercial application to aquaculture production is largely limited to the major finfish and crustacean species (e.g., salmonids, Nile tilapia, shrimp) (Zenger et al., 2019; Lillehammer et al., 2020; Boudry et al., 2021). Studies into the feasibility of applying genomic selection schemes in oyster breeding programmes have shown that for growth (Gutierrez et al., 2018; Vu et al., 2021b), edibility (Vu et al., 2021a), and disease resistance traits (Gutierrez et al., 2020; Vu et al., 2021a), greater genetic gains could be achieved through GS compared to traditional breeding. Nevertheless, the practical application of GS as a selection strategy will likely depend on how cost-effective it is compared to pedigree-based methods. The development of feasible alternatives for reducing genotyping costs, such as using affordable low-density genotyping tools that yield similar accuracies than higher-density panels, will be critical for the potential of GS to be realized by oyster breeding programmes.

In line with the increasing interest in supporting oyster culture and restoration through the breeding of stocks with enhanced performance, the overall aim of this study was to evaluate the potential of GS for the genetic improvement of growth and growth-related (morphometric) traits in the European flat oyster. First, the heritability of total weight, shell length, shell width and shell height was estimated for a hatchery-derived population genotyped using a ~15K SNP array. Second, a genome-wide association (GWAS) analysis was conducted to dissect the genetic architecture of the measured traits. Last, to evaluate whether GS may be an effective and cost-effective strategy for improving traits associated with oyster growth, the accuracy of genomic predictions using reduced density SNP marker panels was assessed.

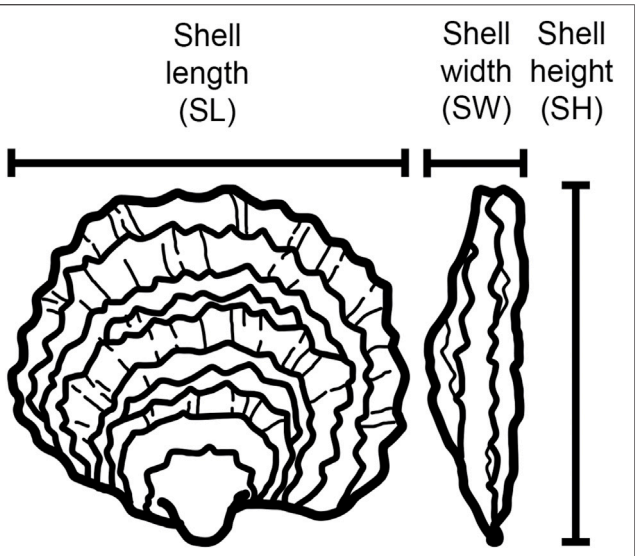



**FIGURE 1 |** Nomenclature of the growth-related morphometric traits measured in this study.

# MATERIALS AND METHODS

## Field Experiment

The European flat oyster population used in this study was generated in a UK-based hatchery (Seasalter Morecombe hatchery) by mass spawning of approximately 40 broodstock parents over several spawning events. The resulting F1 generation was then deployed to Lochnell oysters (56.494°N, 5.459°W) and grown for 6 months in ortac grow-out cages. Next, animals were transferred to the Institute of Marine Sciences at the University of Portsmouth (UK), and maintained in a flow-through system until deployment. During this holding period, ~1,000 randomly selected oysters were individually tagged and their first phenotype measurements recorded (see "Phenotypes" section below). Prior to deployment, animals were cleaned of fouling, washed in fresh water and dried. Embossed plastic tags with unique identifier codes were attached with epoxy resin glue. Animals were returned to aquaria within the hour. Oysters were placed in Aquamesh® cages (L 0.55 m × W 0.55 m × D 0.4 m) at a density of 200 oysters per cage, and deployed 1 m below floating pontoons at Port Hamble Marina (MDL) in the River Hamble (50.861°N, 1.312°W) in January 2019. Mortalities were recorded monthly and dead oysters—i.e., those with empty or gaping shells—were removed from the experiment. General disease status was assessed on subsets of oysters throughout the experiment by histology and *in situ* hybridisation using an adaptation of available methods (Montagnani et al., 2001; Fabioux et al., 2004). In addition, the presence of *Bonamia ostreae*, a protozoan parasite that causes a lethal infection of flat oyster haemocytes (Pichot et al., 1979), was assessed by qPCR following Robert et al. (2009). The prevalence of *B. ostreae* infections was negligible; hence disease status had a minor influence on the assessment of growth traits in the experimental population. After 10 months of growth under field conditions, gill tissue was dissected from individuals alive at the end of the study and preserved in molecular grade absolute ethanol (Fisher Scientific) for genetic analysis.

## Phenotypes

Four growth-associated traits were measured at three time points over the course of 10 months: total weight (TW, the weight of an individual oyster including the shell), shell length (SL, the maximum distance between the anterior and posterior margins), shell height (SH, the maximum distance between the hinge to the furthermost edge), and shell width (SW, the maximum distance at the thickest part of the two shell valves) (**Figure 1**). Weight was recorded in grams up to one decimal place. Shell measurements were taken with traceable digital callipers (Fisher Scientific) with 0.02 mm precision. Oysters were cleaned and defouled before measurements were taken.

## DNA Extraction

Total DNA was isolated from gill tissue following a CTAB (cetyltrimethylammonium bromide)-based extraction protocol [details in Gutierrez et al. (2017)]. The integrity of the extracted DNA was assessed by agarose gel electrophoresis, while DNA purity was verified on a Nanodrop ND-1000 (Thermo Fisher Scientific) spectrophotometer by checking the 260/280 and 260/230 ratios. All samples had 260/280 and 260/230 values ≥ 1.85 and ≥1.96, respectively.

## SNP Genotyping and Quality Control

Whole-genome genotyping of ~15K SNPs was carried out by IdentiGEN (Dublin, Ireland) using the combined-species Affymetrix Axiom oyster SNP-array (Gutierrez et al., 2017). Signal intensity files were imported to the Axiom analysis Suite v4.0.3.3 software for quality control (QC) assessment and genotype calling. Genotypes were generated using the default parameter settings for diploid species, resulting in 11,808 SNPs typed for 870 individuals. To assess the reproducibility of genotype calls, five DNA samples from the same individual were genotyped independently on three different arrays, and their genotype concordance evaluated through an identity-by-state (IBS) analysis. The genotype concordance rate among replicates was 99.7%, demonstrating a high reproducibility of the genotyping assays. The flanking region of these markers were mapped to the *O. edulis* chromosome-level genome assembly (Gundappa et al., 2022). Of the 11,808 SNPs, 10,025 had uniquely mapping probes and were retained for downstream analysis. A total of 1,539 markers (15.4%) were monomorphic in the population under study. Additional QC was performed on markers and samples using Plink v2.0 (Chang et al., 2015). SNP variants were retained for further analysis if they had a call rate > 95% and a minor allele frequency (MAF) > 0.05. These filters removed 4,391 SNPs (leaving a total of 5,634 SNPs), of which the majority were filtered out based on the MAF threshold (i.e., were monomorphic or near-monomorphic in this population). Given that significant sub-clustering was detected in the data (**Supplementary Figure S1**), possibly due to a high variance in the reproductive success of broodstock parents and/or temporal variation in spawning, a k-means clustering method was used to assign individuals into groups. Deviations from Hardy-Weinberg Equilibrium (HWE) were tested separately in each of the three genetic clusters identified by the analysis. SNP markers showing significant deviations (HWE $p$-value < 1e-10) in two of the three clusters were excluded from the analysis. Sample QC included removing individual oysters with a missingness above 5% and high heterozygosity (i.e., more than three median absolute deviations from median). Finally, a principal component analysis (PCA) was performed using a set of ~3.5K SNPs for which no pair of markers within a window of 200 kb had a $r^2 > 0.5$. The top five PCs, which explain 47% of the variance (considering 20 PCs), were fitted in the model to account for the effect of population structure. The final dataset comprised 840 samples genotyped at 4,577 genome-wide SNPs.

## Genetic Parameter Estimation

Genetic parameters for growth-related traits were estimated by fitting the following univariate linear mixed model in GEMMA v0.95alpha (Zhou and Stephens, 2012):

$$y = \mu + \boldsymbol{X}b + \boldsymbol{Z}u + e \quad (1)$$

Where $\boldsymbol{y}$ is the vector of observed phenotypes; $\boldsymbol{\mu}$ is the overall mean of the phenotype in the population; $\mathbf{b}$ is the vector of fixed effects to be fitted (the first five principal components were included as covariates); $\mathbf{u}$ is the vector of the additive genetic effects; $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the corresponding incidence matrices for fixed and additive effects, respectively; and $e$ is a vector of residuals. The following distributions were assumed: $u \sim N(0, G\sigma_u^2)$) and $e \sim N(0, I\sigma_e^2)$. Where $\sigma_u^2$ and $\sigma_e^2$ are the additive genetic and residual variance, respectively, $\mathbf{G}$ is the genomic relationship matrix and $\mathbf{I}$ is the identity matrix. The heritability of growth-related traits was estimated as the ratio of the additive genetic variance to the total phenotypic variance.

Bivariate animal linear models were implemented to estimate the genetic (co)variance between TW, SL, SH and SW. Each bivariate analysis was fitted with the same top 5 PCs mentioned above. Subsequently, genetic correlations among traits were estimated as the ratio of the covariance of two traits to the square root of the product of the variance for each trait. Phenotypic correlations between traits were calculated using the Pearson correlation coefficient.

## Genome-Wide Association Study

To identify SNPs in the flat oyster genome correlated with variation in growth-related traits, a GWAS was performed by implementing the same model described previously in the GEMMA software. SNPs were considered significant at the genome-wide level if their likelihood ratio test P-values surpassed a conservative Bonferroni-corrected significance threshold (α/4,577 = 1.09e-5). To derive a threshold for chromosome-wide (suggestive) significance, α was divided by the average number of SNPs per chromosome (α/457 = 1.09e-4). The single-marker P-values obtained from GEMMA were plotted against their chromosome location using the R package qqman v 0.1.4 (Turner, 2018). To assess the inflation of the association statistics, the genomic control coefficient lambda λGC was calculated following Devlin and Roeder (1999). Candidate genes were searched within 100 kb of the most significant SNP loci using BEDOPS v2.4.26 (Neph et al., 2012).

## Genomic Prediction

To evaluate the accuracy of genomic selection, a 5-fold cross validation approach — animals split into training (80%) and validation (20%) sets—was used on a population of 840 oysters genotyped for 4,577 informative SNP markers. To reduce stochastic effects arising from individual sampling, each analysis was repeated 10 times. For each replicate, animals were randomly partitioned into five subsets (each subset contained 168 individuals). TW, SL, SH and SW phenotypes recorded in individuals allocated to one of the subsets (validation set) were masked. The breeding values of the validation set were then predicted based on the information from the remaining four subsets (training sets) using model (1). The model was fitted using the AIREMLF90 module from BLUPF90. The accuracy of genomic predictions was calculated as follows:

$$Accuracy = \frac{r_{gEBV,\,y}}{h}$$

where $r_{gEBV,y}$ is the correlation between the predicted and the actual phenotypes of the validation set, while $h$ is the square root of the heritability of the trait estimated as described above.

## Evaluation of the Effect of SNP Density on Genomic Predictions

To assess the effect of SNP density on the accuracy of genomic predictions of growth-related traits, SNP panels of varying sizes were randomly sampled from the final pool of QC-filtered array markers ($n$ = 4,577 SNPs). Panels of the following densities were evaluated: 4, 3, 2, 1K, 500, 400, 300, 200 and 100 SNPs. To build the lower-density panels, markers were randomly sampled from the full QC-filtered SNP dataset in proportion to chromosome lengths using the R package CVrepGPAcalc v1.0 (Tsairidou, 2019; Tsairidou et al., 2020). To account for sampling bias, 10 SNP panels were generated for each of the SNP densities. The average genomic prediction accuracies of the different low-density panels were compared against the equivalent accuracy values obtained with the full panel.

# RESULTS AND DISCUSSION

## Growth Traits and Heritability

Improvement of growth rate is typically one of the first traits to be included as a selection target in breeding programmes across many farmed species. In this study, oyster growth rate was assessed in a hatchery-derived oyster population that was translocated to a growing site and monitored for 10 months. The experimental population had a lower genome-wide heterozygosity (Ho = 0.27; He = 0.22) compared to the values reported by Vera et al. (2019) (Ho ≥ 0.31) for a diverse set of flat oyster populations genotyped with the same array. An overall mortality of 14% was observed during the field trial, among which the majority (36%) occurred during a summer month (July). At the end of the experimental period, the *O. edulis* population had the following growth means and standard deviations: +15.7 g (SD = 5.8), 50.8 mm (SD = 7.3), 12.9 mm (SD = 2.3) and 45.8 mm (SD = 8.9), for TW, SH, SW, SL, respectively (**Table 1**). The phenotypic correlation was found to be the highest ($r$ > 0.8) between two pairs of traits: (i) TW and SH, and (ii) TW and SL (**Figure 2**).

For the European oyster population under study, the heritability estimates of these growth-related traits were in the low to moderate range of 0.22 (for SW) to 0.45 (for TW) (**Table 2**). Consistent with similar studies carried out in related oyster species (Xu et al., 2017; Vu et al., 2020), heritability estimates based on SNP markers were higher for total weight than for growth-related morphometric traits (i.e., shell height, shell width and shell length). The estimation of heritability for total weight (TW) was similar to those reported for nine- month-old Portuguese oysters ($h^2$ = 0.45) and a 2-year old Pacific oyster strain ($h^2$ = 0.42) (Xu et al., 2017; Vu et al., 2021b). Total weight, as measured in this study, is a composite phenotype made up of the animal's shell and soft tissue weights, in addition to the weight of any pallial fluid - thus is not a direct reflection of meat yield. Nevertheless, in *C. angulata*, a positive genetic correlation (0.63) has been found between TW and soft tissue weight (Vu et al., 2021b), suggesting that selecting for TW—a trait easier to measure—could lead to improvements in meat yields. Such indirect improvements of correlated traits have been reported in a Portuguese oyster line selected for harvest weight. While the achieved average selection response for total weight at harvest was

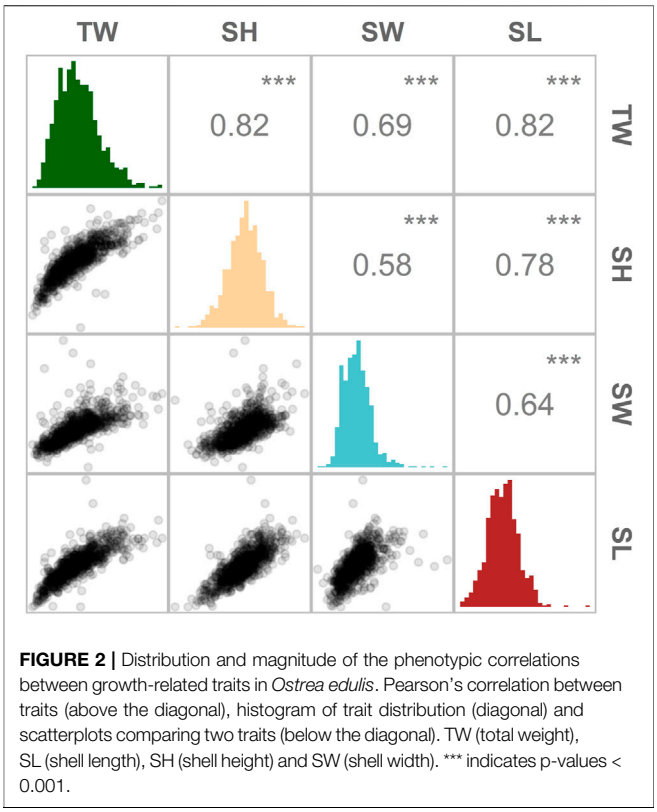



**FIGURE 2 |** Distribution and magnitude of the phenotypic correlations between growth-related traits in *Ostrea edulis*. Pearson's correlation between traits (above the diagonal), histogram of trait distribution (diagonal) and scatterplots comparing two traits (below the diagonal). TW (total weight), SL (shell length), SH (shell height) and SW (shell width). *** indicates p-values < 0.001.

**TABLE 1 |** Summary statistics of the phenotypic data (SD: Standard deviation; CV: coefficient of variation).

| Trait | Unit | Mean | Min | Max | SD | CV (%) |
|---|---|---|---|---|---|---|
| Total weight | g | 15.7 | 4.0 | 38.5 | 5.8 | 36.8 |
| Shell height | mm | 50.8 | 22.9 | 76.1 | 7.3 | 14.4 |
| Shell width | mm | 12.9 | 6.5 | 27.7 | 2.3 | 18.2 |
| Shell length | mm | 45.8 | 22.2 | 94.4 | 8.9 | 19.5 |

**TABLE 2 |** Estimates of heritability ($h^2$) and SE on the diagonal and pairwise genetic correlations (below the diagonal) for growth-related traits in a European flat oyster population.

| Trait | Total weight | Shell height | Shell width | Shell length |
|---|---|---|---|---|
| Total weight | **0.45 (0.06)** | | | |
| Shell height | 0.99 | **0.37 (0.06)** | | |
| Shell width | 0.96 | 0.90 | **0.22 (0.05)** | |
| Shell length | 0.95 | 0.93 | 0.88 | **0.32 (0.06)** |

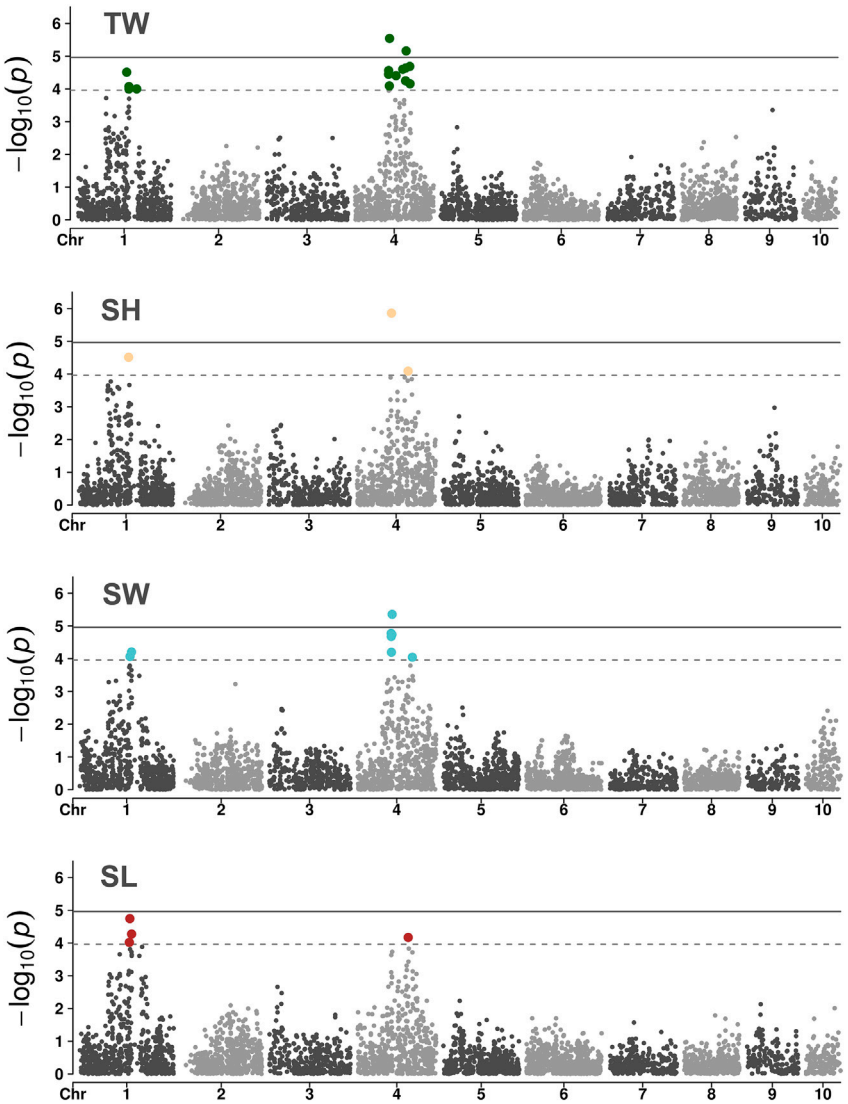



**FIGURE 3 |** Manhattan plots of the GWAS for growth-related traits in a European flat oyster population. Solid lines indicate the threshold value for genome-wide significance. Dashed lines indicate the threshold for a suggestive (chromosome-level) significance.

5.8% per generation, genetic gains were also observed for soft tissue weight, with indirect gains reaching a 1.2% increase per generation (Vu et al., 2020). For the shell-related traits examined in this study (SH, SW and SL), heritability estimates were in line with previous studies (Yuehuan et al., 2017; Gutierrez et al., 2018), and ranged from 0.22 to 0.37. Traditionally, the focus on shell morphometric traits was to improve oyster growth. Nevertheless, in recent years, oyster shell shape is increasingly being viewed as an attractive goal for selective breeding due to its growing importance for consumers (Mizuta and Wikfors, 2019). The perceived attractiveness of an oyster shell can be represented as a secondary trait derived from a ratio between primary (shell dimension) traits, such as the shell width index (Kube et al., 2011). Given that significant heritable variation was observed for the three examined morphometric traits, strategies for homogenizing particular shell shapes may be feasible in *O. edulis*.

## Genome-Wide Association Analysis for Growth-Related Traits

A GWAS of ~4.5K SNPs passing the filtering criteria were genotyped on 840 oysters with phenotypic records to gain insight into the genetic basis of growth rate variation in *O. edulis*. Three of the four examined traits showed association signals surpassing the genome-wide level of significance (**Figure 3**). The genomic inflation factor lambda of the GWAS analysis were close to the desired value ($\lambda$ = 1) (see **Supplementary Figure S2**), indicating that population structure was adequately accounted for by the model. For TW, the GWAS identified two putative quantitative trait loci (QTLs) on chromosome 4 associated with the trait. The presence of two separate QTLs is supported by the low linkage disequilibrium observed between the most significant SNPs at each locus

(pairwise $r^2$ < 0.1). An additional 13 suggestive loci were also identified, of which nine were located in the vicinity of the two abovementioned genome-wide hits and four were found on chromosome 1 (**Supplementary Table S1**). The SNP showing the strongest association with TW (AX-169174635) explained 3% of the phenotypic variance. This lead SNP was also found to be significantly associated with SH and SW. For SL, no SNP reached a genome-wide significance level, although a few of the same markers showing associations with TW, SH and SW surpassed the threshold for suggestive significance. The complete overlap of GWAS hits across the different traits suggests a high degree of shared genetic control among them, consistent with the high positive genetic correlations observed (**Table 2**). Overall, the GWAS results indicate that growth-related traits in *O. edulis* are influenced by many small-effect loci, exhibiting a polygenic architecture, but that two regions on chromosome 4 may have a moderate effect on these traits.

The marker showing the most significant association with TW, SH and SW is located in the exon of a gene annotated as a *N4BP2* (NEDD4 Binding Protein 2)-like protein (Gene ID: *FUN_017843*; Gundappa et al., 2022). The predicted protein product of this gene contains a highly conserved AAA domain, able to bind and hydrolyse ATP (Lupas and Martin, 2002). Proteins with these domains have been shown to be involved in several cell processes, including protein folding, proteolysis and membrane fusion. Further characterization of this *N4BP2*-like protein would help better understand the genetic component of growth variation in oysters. Nevertheless, considering that the candidate marker on *N4BP2* explained a small percentage of the phenotypic variance, independent oyster populations should first be evaluated to confirm the validity of the association signal. A second genome-wide significant association—detected only in the TW GWAS—was located in the exon of an uncharacterized gene (*FUN_018833*) whose product displays > 80% amino acid identity and 99% coverage (best BLASTp hit to NCBI's nr database) with similarly uncharacterized proteins in *C. gigas* and *C. virginica* (NCBI accession numbers XP_011433755.2 and XP_022325737.1, respectively). Additional genes within the two genomic regions (+/- 100 kb) showing significant associations with flat oyster growth traits are shown in **Supplementary Table S2**. Given that the SNPs identified in this study had a small effect on the traits in question, GS would be an effective approach for increasing genetic gains from selection.

## Genomic Selection

The incorporation of genetic markers into breeding programmes requires a previous understanding of the genetic architecture of the targeted trait(s). In the *O. edulis* population under study, the genetic contribution to the observed variation in growth-related traits was largely polygenic in nature. For the improvement of polygenic traits, genomic selection has been shown to be superior to alternative marker-aided selection due to genome-wide markers capturing a higher proportion of the genetic variation in a trait compared to individual QTL-targeted markers. Consequently, by means of applying GS, higher predictions have been achieved for several production traits in a wide range of commercially important aquaculture species [reviewed in Houston et al. (2020)]. Despite GS not yet being widely operational in oyster breeding programmes

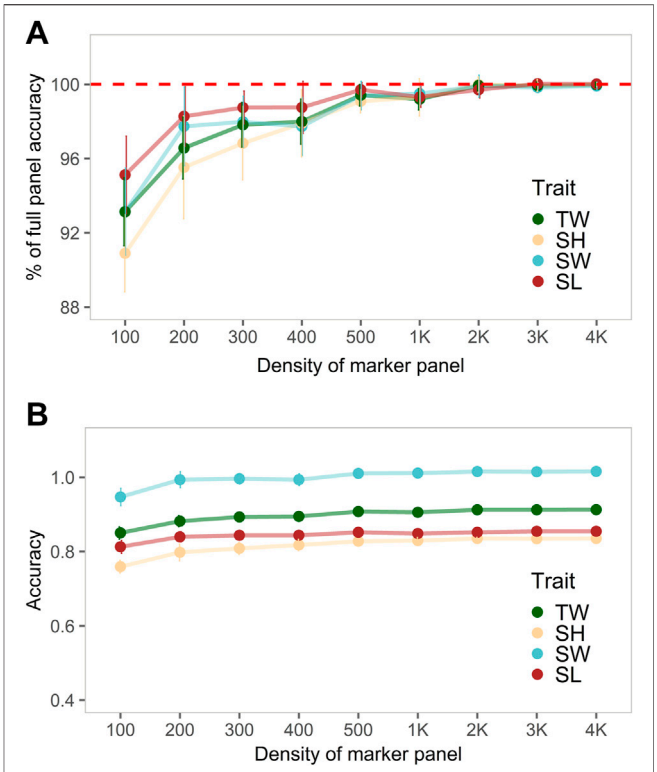



**FIGURE 4 |** Evaluation of the effect of SNP density on genomic predictions of growth related traits in a European flat oyster population. **(A)** Percentage of the maximum genomic prediction accuracy achieved using different lower density SNP panels. Values were calculated by dividing the mean accuracy (averaged over ten replicates) estimated at each nominal SNP density by the accuracy obtained using the full SNP dataset. **(B)** Average genomic prediction accuracy values obtained for oyster growth traits at different panel densities.

(Boudry et al., 2021), studies have demonstrated the potential of incorporating genome-wide information into selection schemes in these taxa. In the Pacific oyster, Gutierrez et al. (2018) showed that prediction accuracies for growth-traits increased ~25–30% when the genetic merit of individuals was estimated from SNP markers using the Genomic Best Linear Unbiased Prediction (GBLUP) model (VanRaden, 2008) compared to a classical pedigree-based approach (PBLUP). Similar results were reported in the Portuguese oyster, as prediction accuracies increased 7–42% for growth-related traits when EBVs obtained by GBLUP were compared to those obtained by PBLUP (Vu et al., 2021a). Since the flat oyster population under study derived from a mass-spawning event, the pedigree structure was unknown. Therefore, comparisons between pedigree and genome based methods for estimating breeding values (e.g., GBLUP and Bayesian approaches) could not be performed.

One of the major barriers of implementing GS is the high number of markers required to accurately predict EBVs and the cost of genotyping these markers (Goddard and Hayes, 2007). Therefore, the design of a strategy to reduce the cost of genotyping is critical for the extensive adoption of genomic prediction approaches in aquaculture breeding programmes.

One such strategy involves genotyping the minimum number of markers required to achieve maximal accuracy, which by definition is equal to that obtained with a full panel of markers. As shown by Kriaridou et al. (2020) for different aquaculture species, the use of low-density SNP panels has the potential to achieve similar EBV accuracies as when using medium density genotype datasets of around 7–14K SNPs. The authors estimated that only 1,000 to 2,000 SNPs are required to achieve maximal accuracy. These results were shown to be consistent across a range of traits (e.g., disease resistance, growth) and species (e.g., Pacific oyster, Atlantic salmon) showing robustness to differences in family structure, genotyping approach, trait heritability and the underlying genetic architecture. In agreement with these findings, maximal accuracy was attained herein for all the assessed growth-related traits at a minimum density of 2K SNPs, with only a slight decline in accuracy observed at the lower densities evaluated (**Figure 4A**). Consequently, a reduction in the costs of applying GS for improving growth traits in *O. edulis* can be achieved by means of exploiting low-density SNP panels. Although low-density panels might not accurately capture the genetic resemblance among individuals within a population, and therefore show reduced genetic variance estimations and EBV predictions when compared with high density panels, their use has been widely evaluated and suggested for different aquaculture species and traits. Furthermore, studies have shown that low-density panels can achieve higher accuracies than the classical pedigree-based approach, being a feasible alternative to identify candidates with the highest genetic merit. For example, in rainbow trout (*Oncorhynchus mykiss*) Vallejo et al. (2018) showed that at least 200 SNPs could exceed PBLUP accuracies for bacterial cold water disease resistance. Whilst Al-Tobasei et al. (2021) found a similar trend when using between 500–1000 SNPs for fillet yield traits. To date, the utilization of low-density panels to decrease the cost of genomic evaluations has also been tested in several aquaculture species, including Atlantic salmon (*Salmo salar*) (Tsai et al., 2016; Correa et al., 2017), rainbow trout (Yoshida et al., 2018; Al-Tobasei et al., 2021), and Nile tilapia (*Oreochromis niloticus*) (Yoshida et al., 2019; Barría et al., 2021), suggesting that the development of cost-effective strategies for applying GS will be key for shaping modern aquaculture breeding programs.

Although our results highlight the possibility of reducing the genotyping costs associated with genomic prediction approaches, caution should be taken as even for the smallest marker density (i.e., 100 SNPs), prediction accuracies (averaged over ten replicates) were high and close to the value obtained with the full marker panel. By using only 100 SNPs the estimated decrease in the accuracy of genomic breeding values (GEBVs) was of 5% for SL, 7% for TW and SW, and 10% for SH (**Figure 4A**). These values highly exceed those reported in the literature for aquaculture species, where reductions > 20% were estimated for panels with 100 SNPs compared to a complete dataset (Kriaridou et al., 2020). The relative stability of GEBVs observed across different marker densities (**Figure 4B**) is likely explained by the underlying genetic structure of the dataset. For this study, 40 potential parents were placed in the same tank and spawned during successive events. The genetic analysis of the progeny revealed that the population was dominated (70% of the sample size; $n$ = 589) by a group of highly related individuals (**Supplementary Figures S1, S3**), suggesting there was a large variance in reproductive success among breeders, as previously reported in mass spawning of oysters (Lallias et al., 2010). In the context of GS, the inclusion of highly related animals in the training and validation sets results in only a small number of markers being required to capture the haplotype effects, as related animals share longer haplotypes (Hickey et al., 2014). The fact that in the current study animals grouped in the reference and validation data sets were highly related would have likely increased the accuracy of predicted GEBVs, even when animals are genotyped at low density, as also shown by Fraslin et al. (2022) in Atlantic salmon. Therefore, additional populations with different effective population sizes, genetic backgrounds, and degrees of relatedness should be assessed to obtain estimates to be expected in practical breeding programs. Future work focused on evaluating the extent to which low-density panels and alternative strategies (e.g., genotype imputation) can be used to reduce genotyping costs will be key for the cost-effective exploitation of GS by oyster breeding programmes.

## CONCLUSION

Growth-related traits in *O. edulis* had moderate to low heritability estimates, ranging from 0.22 (for SW) to 0.45 (for TW). High genetic correlations were identified between all traits (> 0.9); hence, TW—a trait easier to measure—can potentially be used as a proxy phenotype for improving the three examined morphometric traits (SH, SW and SL). The GWAS results revealed that growth traits were largely polygenic, but with two distinct QTLs on chromosome 4 reaching genome-wide significance. Prediction accuracies were high for all traits (> 0.83), with minimal differences observed when comparing estimates obtained using different marker densities. Altogether, these results suggest that the high prediction accuracies found in this study could have been influenced by the highly unbalanced family structure of the experimental population. Consequently, although low-density SNP panels appear as a promising cost-effective GS strategy, additional populations with different degrees of genetic relatedness should be assessed to derive estimates of prediction accuracies to be expected in practical breeding programmes in oysters.

## DATA AVAILABILITY STATEMENT

The data presented in this study are available in the Mendeley Data repository as independent genotype and phenotype datasets accessible at https://doi.org/10.17632/v7ppb2xr99.2 and https://doi.org/10.17632/sdtjyys7gr.1, respectively.

## AUTHOR CONTRIBUTIONS

TB and RH conceptualized the study. JH provided the animals for the study. CP, CH, DM, TB, LH, JKH, JP and JN performed the

experiment including sample collection and recording phenotype information. AP extracted the DNA used for genotyping. CP and AB performed the analysis. MG, DM and CH evaluated disease status. MKG and DJM provided intellectual input to the analysis. CP and AB wrote the manuscript with input from all authors.

## FUNDING

The authors acknowledge funding from the Biotechnology and Biological Sciences Research Council (BBSRC), including Institute Strategic Programme grants (BBS/E/D/20002172, BBS/E/D/30002275 and BBS/E/D/10002070), a grant within the AquaLeap project (BB/S004181/1), funding from Sustainable Aquaculture Innovation Centre (SAIC), Blue Marine Foundation and National Fish and Wildlife Foundation (NFWF).

## ACKNOWLEDGMENTS

The authors would like to acknowledge MDL Port Hamble Marina for allowing positioning of oyster cages within the Marina, and thank Eric Harris-Scott, Matthew Sanders, Monica Fabra, Tim Regan and Zenaba Khatir for their invaluable help with setup and sampling of the field experiment. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.926638/full#supplementary-material

**Supplementary Figure S1 |** PCA of the *O.edulis* population under study showing the dominance of single cluster comprised of highly related individuals (clst_0).

**Supplementary Figure S2 |** Quantile-quantile (Q-Q) plots showing the distribution of the expected (red dashed line) versus observed P-values of the GWAS of growth-related traits.

**Supplementary Figure S3 |** Example of the pairwise kinship coefficient between individuals in the training and validation set across randomly selected iterations of different 5-fold cross validation (CV) replicates (ten replicates in total). Boxplots show the distribution of values for pairs of individuals (one from the training set and one from the validation set) belonging to the same or different population clusters. Relatedness was inferred using the KING-robust method implemented in Plink v2.0.

## REFERENCES

Allen, S. K., Gaffney, P. M., and Ewart, J. W. (1993). *Genetic Improvement of the Eastern Oyster for Growth and Disease Resistance in the Northeast*. Massachusetts: Northeastern Regional Aquaculture Center, 210.

Al-Tobasei, R., Ali, A., Garcia, A. L. S., Lourenco, D., Leeds, T., and Salem, M. (2021). Genomic Predictions for Fillet Yield and Firmness in Rainbow Trout Using Reduced-Density SNP Panels. *BMC Genomics* 22, 92. doi:10.1186/s12864-021-07404-9

Barría, A., Benzie, J. A. H., Houston, R. D., De Koning, D.-J., and de Verdal, H. (2021). Genomic Selection and Genome-wide Association Study for Feed-Efficiency Traits in a Farmed Nile Tilapia (*Oreochromis niloticus*) Population. *Front. Genet.* 12, 737906. doi:10.3389/fgene.2021.737906

Boeliguf, G. (2000). Present Status of the French Aquaculture. *Aquac. Sci.* 48 (2), 243–248.

Botta, R., Asche, F., Borsum, J. S., and Camp, E. V. (2020). A Review of Global Oyster Aquaculture Production and Consumption. *Mar. Policy* 117, 103952. doi:10.1016/j.marpol.2020.103952

Boudry, P., Allal, F., Aslam, M. L., Bargelloni, L., Bean, T. P., Brard-Fudulea, S., et al. (2021). Current Status and Potential of Genomic Selection to Improve Selective Breeding in the Main Aquaculture Species of International Council for the Exploration of the Sea (ICES) Member Countries. *Aquac. Rep.* 20, 100700. doi:10.1016/j.aqrep.2021.100700

Boutet, I., Monteiro, H. J. A., Takeuchi, T., Baudry, L., Bonnivard, E., Billoud, B., et al. (2022). Chromosomal Assembly of the Flat Oyster (*Ostrea Edulis* L.) Genome as a New Genetic Resource for Aquaculture. *bioRxiv [Preprint]*. Available at: https://doi.org/10.1101/2022.06.26.497643. (Accessed June 29, 2022).

Camara, M. D., Yen, S., Kaspar, H. F., Kesarcodi-Watson, A., King, N., Jeffs, A. G., et al. (2017). Assessment of Heat Shock and Laboratory Virus Challenges to Selectively Breed for Ostreid Herpesvirus 1 (OsHV-1) Resistance in the Pacific Oyster, *Crassostrea gigas*. *Aquaculture* 469, 50–58. doi:10.1016/j.aquaculture.2016.11.031

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the Challenge of Larger and Richer Datasets. *GigaSci* 4, 7. doi:10.1186/s13742-015-0047-8

Colsoul, B., Boudry, P., Pérez-Parallé, M. L., Bratoš Cetinić, A., Hugh-Jones, T., Arzul, I., et al. (2021). Sustainable Large-scale Production of European Flat Oyster ( *Ostrea edulis*) Seed for Ecological Restoration and Aquaculture: A Review. *Rev. Aquac.* 13, 1423–1468. doi:10.1111/raq.12529

Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P., and Yáñez, J. M. (2017). The Use of Genomic Information Increases the Accuracy of Breeding Value Predictions for Sea Louse (*Caligus rogercresseyi*) Resistance in Atlantic Salmon (*Salmo salar*). *Genet. Sel. Evol.* 49, 15. doi:10.1186/s12711-017-0291-8

Culloty, S. C., Cronin, M. A., and Mulcahy, M. F. (2004). Potential Resistance of a Number of Populations of the Oyster *Ostrea edulis* to the Parasite *Bonamia ostreae*. *Aquaculture* 237, 41–58. doi:10.1016/j.aquaculture.2004.04.007

Dégremont, L., Nourry, M., and Maurouard, E. (2015). Mass Selection for Survival and Resistance to OsHV-1 Infection in *Crassostrea gigas* Spat in Field Conditions: Response to Selection after Four Generations. *Aquaculture* 446, 111–121. doi:10.1016/j.aquaculture.2015.04.029

De Melo, C. M. R., Durland, E., and Langdon, C. (2016). Improvements in Desirable Traits of the Pacific Oyster, *Crassostrea gigas*, as a Result of Five Generations of Selection on the West Coast, USA. *Aquaculture* 460, 105–115. doi:10.1016/j.aquaculture.2016.04.017

Devlin, B., and Roeder, K. (1999). Genomic Control for Association Studies. *Biometrics* 55, 997–1004. doi:10.1111/j.0006-341x.1999.00997.x

Evans, S., and Langdon, C. (2006). Direct and indirect responses to selection on individual body weight in the Pacific oyster (*Crassostrea gigas*). *Aquaculture* 261, 522–534. doi:10.1016/j.aquaculture.2006.07.037

Evans, F., Matson, S., Brake, J., and Langdon, C. (2004). The Effects of Inbreeding on Performance Traits of Adult Pacific Oysters (*Crassostrea gigas*). *Aquaculture* 230, 89–98. doi:10.1016/j.aquaculture.2003.09.023

Fabioux, C., Huvet, A., Lelong, C., Robert, R., Pouvreau, S., Daniel, J. Y., et al. (2004). Oyster Vasa-like Gene as a Marker of the Germline Cell Development in *Crassostrea gigas*. *Biochem. Biophys. Res. Commun.* 320, 592–598. doi:10.1016/j.bbrc.2004.06.009

FAO (2022). *Cultured Aquatic Species Information Programme (CASIP)* Ostrea Edulis *Linnaeus 1758*. https://www.fao.org/fishery/en/culturedspecies/ostrea_edulis/en. (Accessed June 26, 2022).

Fitzer, S. C., McGill, R. A. R., Torres Gabarda, S., Hughes, B., Dove, M., O'Connor, W., et al. (2019). Selectively Bred Oysters Can Alter Their Biomineralization Pathways, Promoting Resilience to Environmental Acidification. *Glob. Change Biol.* 25, 4105–4115. doi:10.1111/gcb.14818

Fraslin, C., Yáñez, J. M., Robledo, D., and Houston, R. D. (2022). The Impact of Genetic Relationship between Training and Validation Populations on

Genomic Prediction Accuracy in Atlantic Salmon. *Aquac. Rep.* 23, 101033. doi:10.1016/j.aqrep.2022.101033

Goddard, M. E., and Hayes, B. J. (2007). Genomic Selection. *J. Anim. Breed. Genet.* 124, 323–330. doi:10.1111/j.1439-0388.2007.00702.x

Grizel, H., and Héral, M. (1991). Introduction into France of the Japanese Oyster (*Crassostrea gigas*). *ICES J. Mar. Sci.* 47, 399–403. doi:10.1093/icesjms/47.3.399

Grizzle, R. E., Ward, K. M., Peter, C. R., Cantwell, M., Katz, D., and Sullivan, J. (2017). Growth, Morphometrics and Nutrient Content of Farmed Eastern Oysters, *Crassostrea Virginica* (Gmelin), in New Hampshire, USA. *Aquac. Res.* 48, 1525–1537. doi:10.1111/are.12988

Gundappa, M. K., Peñaloza, C., Regan, T., Boutet, I., Tanguy, A., Houston, R. D., et al. (2022). A Chromosome Level Reference Genome for European Flat Oyster (*Ostrea Edulis* L.) *bioRxiv [Preprint]*. Available at: https://doi.org/10.1101/2022.06.26.497633. (Accessed June 29, 2022)

Gutierrez, A. P., Turner, F., Gharbi, K., Talbot, R., Lowe, N. R., Peñaloza, C., et al. (2017). Development of a Medium Density Combined-Species SNP Array for Pacific and European Oysters (*Crassostrea gigas* and *Ostrea edulis*). *G3 (Bethesda)* 7, 2209–2218. doi:10.1534/g3.117.041780

Gutierrez, A. P., Matika, O., Bean, T. P., and Houston, R. D. (2018). Genomic Selection for Growth Traits in Pacific Oyster (*Crassostrea gigas*): Potential of Low-Density Marker Panels for Breeding Value Prediction. *Front. Genet.* 9, 391. doi:10.3389/fgene.2018.00391

Gutierrez, A. P., Symonds, J., King, N., Steiner, K., Bean, T. P., and Houston, R. D. (2020). Potential of Genomic Selection for Improvement of Resistance to Ostreid Herpesvirus in Pacific Oyster ( *Crassostrea gigas* ). *Anim. Genet.* 51, 249–257. doi:10.1111/age.12909

He, X., Li, C., Qi, H., Meng, J., Wang, W., Wu, F., 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. doi:10.1016/j.aquaculture.2021.736926

He, X., Wu, F., Qi, H., Meng, J., Wang, W., Liu, M., et al. (2022). Whole-genome Resequencing Reveals the Single Nucleotide Polymorphisms Associated with Shell Shape in *Crassostrea gigas*. *Aquaculture* 547, 737502. doi:10.1016/j.aquaculture.2021.737502

Helmer, L., Farrell, P., Hendy, I., Harding, S., Robertson, M., and Preston, J. (2019). Active Management is Required to Turn the Tide for Depleted *Ostrea edulis* Stocks from the Effects of Overfishing, Disease and Invasive Species. *PeerJ* 7, e6431. doi:10.7717/peerj.6431

Hickey, J. M., Dreisigacker, S., Crossa, J., Hearne, S., Babu, R., Prasanna, B. M., et al. (2014). Evaluation of Genomic Selection Training Population Designs and Genotyping Strategies in Plant Breeding Programs Using Simulation. *Crop Sci.* 54, 1476–1488. doi:10.2135/cropsci2013.03.0195

Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., et al. (2020). Harnessing Genomics to Fast-Track Genetic Improvement in Aquaculture. *Nat. Rev. Genet.* 21, 389–409. doi:10.1038/s41576-020-0227-y

Jones, D. B., Jerry, D. R., Khatkar, M. S., Raadsma, H. W., and Zenger, K. R. (2013). A High-Density SNP Genetic Linkage Map for the Silver-Lipped Pearl Oyster, *Pinctada maxima*: A Valuable Resource for Gene Localisation and Marker-Assisted Selection. *BMC Genomics* 14, 810. doi:10.1186/1471-2164-14-810

Jones, D. B., Jerry, D. R., Khatkar, M. S., Moser, G., Raadsma, H. W., Taylor, J. J., et al. (2014). Determining Genetic Contributions to Host Oyster Shell Growth: Quantitative Trait Loci and Genetic Association Analysis for the Silver-Lipped Pearl Oyster, *Pinctada maxima*. *Aquaculture* 434, 367–375. doi:10.1016/j.aquaculture.2014.08.040

Korringa, P. (1976). *Farming the Flat Oyster of the Genus Ostrea*. Amsterdam: ElsevierScientific.

Kriaridou, C., Tsairidou, S., Houston, R. D., and Robledo, D. (2020). Genomic Prediction Using Low Density Marker Panels in Aquaculture: Performance across Species, Traits, and Genotyping Platforms. *Front. Genet.* 11, 124. doi:10.3389/fgene.2020.00124

Kube, P., Cunningham, M., Dominik, S., Parkinson, S., Finn, B., Henshall, J., et al. (2011). *Enhancement of the Pacific Oyster Selective Breeding Program*. Australia: FRDC and Seafood CRC.

Lallias, D., Taris, N., Boudry, P., Bonhomme, F., and Lapègue, S. (2010). Variance in the Reproductive Success of Flat Oyster *Ostrea edulis* L. Assessed by Parentage Analyses in Natural and Experimental Conditions. *Genet. Res.* 92, 175–187. doi:10.1017/S0016672310000248

Lapegue, S., Harrang, E., Heurtebise, S., Flahauw, E., Donnadieu, C., Gayral, P., et al. (2014). Development of SNP-Genotyping Arrays in Two Shellfish Species. *Mol. Ecol. Resour.* 14, 820–830. doi:10.1111/1755-0998.12230

Li, C., Wang, J., Song, K., Meng, J., Xu, F., Li, L., et al. (2018). Construction of a High-Density Genetic Map and Fine QTL Mapping for Growth and Nutritional Traits of *Crassostrea gigas*. *BMC Genomics* 19, 626. doi:10.1186/s12864-018-4996-z

Li, A., Dai, H., Guo, X., Zhang, Z., Zhang, K., Wang, C., et al. (2021). Genome of the Estuarine Oyster Provides Insights into Climate Impact and Adaptive Plasticity. *Commun. Biol.* 4, 1287. doi:10.1038/s42003-021-02823-6

Lillehammer, M., Bangera, R., Salazar, M., Vela, S., Erazo, E. C., Suarez, A., et al. (2020). Genomic Selection for White Spot Syndrome Virus Resistance in Whiteleg Shrimp Boosts Survival under an Experimental Challenge Test. *Sci. Rep.* 10, 20571. doi:10.1038/s41598-020-77580-3

Liu, S., Li, L., Zhang, S., Wang, W., Yang, J., and Zhang, G. (2019). Heritability Estimates for Nutritional Quality-related Traits of the Pacific Oyster, *Crassostrea gigas*. *J. World Aquacult. Soc.* 50, 738–748. doi:10.1111/jwas.12588

Lupas, A., and Martin, J. (2002). AAA Proteins. *Curr. Opin. Struct. Biol.* 12, 746–753. doi:10.1016/s0959-440x(02)00388-3

McCarty, A. J., Allen, S. K., and Plough, L. V. (2022). Genome-wide Analysis of Acute Low Salinity Tolerance in the Eastern Oyster *Crassostrea virginica* and Potential of Genomic Selection for Trait Improvement. *G3 (Bethesda)* 12, jkab368. doi:10.1093/g3journal/jkab368

Meng, J., Song, K., Li, C., Liu, S., Shi, R., Li, B., et al. (2019). Genome-wide Association Analysis of Nutrient Traits in the Oyster *Crassostrea gigas*: Genetic Effect and Interaction Network. *BMC Genomics* 20, 625. doi:10.1186/s12864-019-5971-z

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-wide Dense Marker Maps. *Genetics* 157, 1819–1829. doi:10.1093/genetics/157.4.1819

Mizuta, D. D., and Wikfors, G. H. (2019). Seeking the Perfect Oyster Shell: A Brief Review of Current Knowledge. *Rev. Aquacult.* 11, 586–602. doi:10.1111/raq.12247

Modak, T. H., Literman, R., Puritz, J. B., Johnson, K. M., Roberts, E. M., Proestou, D., et al. (2021). Extensive Genome-wide Duplications in the Eastern Oyster (*Crassostrea virginica*). *Phil. Trans. R. Soc. B* 376, 20200164. doi:10.1098/rstb.2020.0164

Montagnani, C., Le Roux, F., Berthe, F., and Escoubas, J.-M. (2001). Cg-TIMP, an Inducible Tissue Inhibitor of Metalloproteinase from the Pacific oyster *Crassostrea gigas* with a Potential Role in Wound Healing and Defense Mechanisms[1]. *FEBS Lett.* 500, 64–70. doi:10.1016/s0014-5793(01)02559-5

Naciri-Graven, Y., Martin, A.-G., Baud, J.-P., Renault, T., and Gérard, A. (1998). Selecting the Flat Oyster *Ostrea edulis* (L.) for Survival when Infected with the Parasite *Bonamia ostreae*. *J. Exp. Mar. Biol. Ecol.* 224, 91–107. doi:10.1016/S0022-0981(97)00171-8

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., et al. (2012). BEDOPS: High-Performance Genomic Feature Operations. *Bioinformatics* 28, 1919–1920. doi:10.1093/bioinformatics/bts277

Newkirk, G. F., and Haley, L. E. (1982). Progess in Selection for Growth Rate in the European Oyster *Ostrea edulis*. *Mar. Ecol. Prog. Ser.* 10, 3. doi:10.3354/meps010077

Newkirk, G. F., and Haley, L. E. (1983). Selection for Growth Rate in the European Oyster, *Ostrea edulis*: Response of Second Generation Groups. *Aquaculture* 33, 149–155. doi:10.1016/0044-8486(83)90396-4

Peñaloza, C., Gutierrez, A. P., Eöry, L., Wang, S., Guo, X., Archibald, A. L., et al. (2021). A Chromosome-Level Genome Assembly for the Pacific Oyster *Crassostrea gigas*. *GigaScience* 10, giab020. doi:10.1093/gigascience/giab020

Pichot, Y., Comps, M., Tige, G., Grizel, H., and Rabouin, M. A. (1979). Recherches sur Bonamia ostreae gen. n., sp. n., parasite nouveau de l' huitre plate *Ostrea edulis* L. [France]. *Rev. Trav. Inst. Peches Marit.* 43, 131–140.

Plough, L. V. (2016). Genetic Load in Marine Animals: a Review. *Curr. Zool.* 62, 567–579. doi:10.1093/cz/zow096

Pogoda, B., Boudry, P., Bromley, C., Cameron, T. C., Colsoul, B., Donnan, D., et al. (2020). NORA Moving Forward: Developing an Oyster Restoration Network in Europe to Support the Berlin Oyster Recommendation. *Aquat. Conserv. Mar. Freshw. Ecosyst.* 30, 2031–2037. doi:10.1002/aqc.3447

Pogoda, B. (2019). Current Status of European Oyster Decline and Restoration in Germany. *Humanities* 8, 9. doi:10.3390/h8010009

Preston, J., Fabra, M., Helmer, L., Johnson, E., Harris-Scott, E., and Hendy, I. W. (2020). Interactions of Larval Dynamics and Substrate Preference Have Ecological Significance for Benthic Biodiversity and *Ostrea edulis* Linnaeus, 1758 in the Presence of Crepidula Fornicata. *Aquat. Conserv. Mar. Freshw. Ecosyst.* 30, 2133–2149. doi:10.1002/aqc.3446

Proestou, D. A., Vinyard, B. T., Corbett, R. J., Piesz, J., Allen, S. K., Small, J. M., et al. (2016). Performance of Selectively-Bred Lines of Eastern Oyster, *Crassostrea virginica*, across Eastern US Estuaries. *Aquaculture* 464, 17–27. doi:10.1016/j.aquaculture.2016.06.012

Qi, H., Song, K., Li, C., Wang, W., Li, B., Li, L., et al. (2017). Construction and Evaluation of a High-Density SNP Array for the Pacific Oyster (*Crassostrea gigas*). *PLoS One* 12, e0174007. doi:10.1371/journal.pone.0174007

Qi, H., Li, L., and Zhang, G. (2021). Construction of a Chromosome-level Genome and Variation Map for the Pacific Oyster *Crassostrea gigas*. *Mol. Ecol. Resour.* 21, 1670–1685. doi:10.1111/1755-0998.13368

Ragone Calvo, L. M., Calvo, G. W., and Burreson, E. M. (2003). Dual Disease Resistance in a Selectively Bred Eastern Oyster, *Crassostrea virginica*, Strain Tested in Chesapeake Bay. *Aquaculture* 220, 69–87. doi:10.1016/S0044-8486(02)00399-X

Renault, T., Cochennec, N., and Grizel, H. (1995). Bonamia Ostreae, Parasite of the European Flat Oyster, *Ostrea edulis*, Does Not Experimentally Infect the Japanese Oyster, *Crassostrea Gigas*. *Bull. Eur. Ass. Fish. Pathol.* 15, 78.

Robert, R., Borel, M., Pichot, Y., and Trut, G. (1991). Growth and Mortality of the European oyster *Ostrea edulis* in the Bay of Arcachon (France). *Aquat. Living Resour.* 4, 265–274. doi:10.1051/alr:1991028

Robert, M., Garcia, C., Chollet, B., Lopez-Flores, I., Ferrand, S., François, C., et al. (2009). Molecular Detection and Quantification of the Protozoan *Bonamia ostreae* in the Flat Oyster, *Ostrea edulis*. *Mol. Cell. Probes* 23, 264–271. doi:10.1016/j.mcp.2009.06.002

Sas, H., Deden, B., Kamermans, P., zu Ermgassen, P. S. E., Pogoda, B., Preston, J., et al. (2020). Bonamia Infection in Native Oysters (*Ostrea edulis*) in Relation to European Restoration Projects. *Aquat. Conserv. Mar. Freshw. Ecosyst.* 30, 2150–2162. doi:10.1002/aqc.3430

Thurstan, R. H., Hawkins, J. P., Raby, L., and Roberts, C. M. (2013). Oyster (*Ostrea edulis*) Extirpation and Ecosystem Transformation in the Firth of Forth, Scotland. *J. Nat. Conserv.* 21, 253–261. doi:10.1016/j.jnc.2013.01.004

Toro, J., and Newkirk, G. (1990). Divergent Selection for Growth Rate in the European Oyster *Ostrea edulis* Response to Selection and Estimation of Genetic Parameters. *Mar. Ecol. Prog. Ser.* 62, 219–227. doi:10.3354/meps062219

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Bron, J. E., Taggart, J. B., et al. (2016). Genomic Prediction of Host Resistance to Sea Lice in Farmed Atlantic Salmon Populations. *Genet. Sel. Evol.* 48, 47. doi:10.1186/s12711-016-0226-9

Tsairidou, S., Hamilton, A., Robledo, D., Bron, J. E., and Houston, R. D. (2020). Optimizing Low-Cost Genotyping and Imputation Strategies for Genomic Selection in Atlantic Salmon. *G3 (Bethesda)* 10, 581–590. doi:10.1534/g3.119.400800

Tsairidou, S. (2019). CVrepGPAcalc. Available at: https://github.com/SmaragdaT/CVrep/tree/master/CVrepGPAcalc (Accessed March 15, 2020).

Turner, S. D. (2018). Qqman: An R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots. *J. Open Source Softw.* 3 (25), 731. doi:10.21105/joss.00731

Vallejo, R. L., Silva, R. M. O., Evenhuis, J. P., Gao, G., Liu, S., Parsons, J. E., et al. (2018). Accurate Genomic Predictions for BCWD Resistance in Rainbow Trout are Achieved Using Low-density SNP Panels: Evidence that Long-range LD is a Major Contributing Factor. *J. Anim. Breed. Genet.* 135, 263–274. doi:10.1111/jbg.12335

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91, 4414–4423. doi:10.3168/jds.2007-0980

Varney, R. L., and Wilbur, A. E. (2020). Analysis of Genetic Variation and Inbreeding Among Three Lines of Hatchery-Reared *Crassostrea virginica* Broodstock. *Aquaculture* 527, 735452. doi:10.1016/j.aquaculture.2020.735452

Vera, M., Pardo, B. G., Cao, A., Vilas, R., Fernández, C., Blanco, A., et al. (2019). Signatures of Selection for Bonamiosis Resistance in European Flat Oyster (*Ostrea edulis*): New Genomic Tools for Breeding Programs and Management of Natural Resources. *Evol. Appl.* 12 (9), 1781–1796. doi:10.1111/eva.12832

Vu, S. V., Knibb, W., Nguyen, N. T. H., Vu, I. V., O'Connor, W., Dove, M., et al. (2020). First Breeding Program of the Portuguese Oyster *Crassostrea angulata* Demonstrated Significant Selection Response in Traits of Economic Importance. *Aquaculture* 518, 734664. doi:10.1016/j.aquaculture.2019.734664

Vu, S. V., Gondro, C., Nguyen, N. T. H., Gilmour, A. R., Tearle, R., Knibb, W., et al. (2021a). Prediction Accuracies of Genomic Selection for Nine Commercially Important Traits in the Portuguese Oyster (*Crassostrea angulata*) Using DArT-Seq Technology. *Genes* 12, 210. doi:10.3390/genes12020210

Vu, S. V., Knibb, W., Gondro, C., Subramanian, S., Nguyen, N. T. H., Alam, M., et al. (2021b). Genomic Prediction for Whole Weight, Body Shape, Meat Yield, and Color Traits in the Portuguese Oyster *Crassostrea angulata*. *Front. Genet.* 12, 661276. doi:10.3389/fgene.2021.661276

Walne, P. R., and Helm, M. M. (1979). *Introduction of* Crassostrea gigas *into the United Kingdom*. Cambridge: MIT Press, 83–105.

Wan, S., Li, Q., Yu, H., Liu, S., and Kong, L. (2020). Estimating Heritability for Meat Composition Traits in the Golden Shell Strain of Pacific Oyster (*Crassostrea gigas*). *Aquaculture* 516, 734532. doi:10.1016/j.aquaculture.2019.734532

Wang, J., Li, L., and Zhang, G. (2016). A High-Density SNP Genetic Linkage Map and QTL Analysis of Growth-Related Traits in a Hybrid Family of Oysters (*Crassostrea gigas* × *Crassostrea angulata*) Using Genotyping-By-Sequencing. *G3 (Bethesda)* 6, 1417–1426. doi:10.1534/g3.116.026971

Ward, R. D., Thompson, P. A., Appleyard, S. A., Swan, A. A., and Kube, P. D. (2005). Sustainable Genetic Improvement of Pacific Oysters in Tasmania and South Australia. FRDC Project 2000/206. CSIRO Marine and Atmospheric Research.

Xu, L., Li, Q., Yu, H., and Kong, L. (2017). Estimates of Heritability for Growth and Shell Color Traits and Their Genetic Correlations in the Black Shell Strain of Pacific Oyster *Crassostrea gigas*. *Mar. Biotechnol.* 19, 421–429. doi:10.1007/s10126-017-9772-6

Yang, B., Zhai, S., Zhang, F., Wang, H., Ren, L., Li, Y., et al. (2022). Genome-wide Association Study toward Efficient Selection Breeding of Resistance to *Vibrio alginolyticus* in Pacific Oyster, *Crassostrea gigas*. *Aquaculture* 548, 737592. doi:10.1016/j.aquaculture.2021.737592

Yin, X., Arias-Pérez, A., Kitapci, T. H., and Hedgecock, D. (2020). High-Density Linkage Maps Based on Genotyping-By-Sequencing (GBS) Confirm a Chromosome-Level Genome Assembly and Reveal Variation in Recombination Rate for the Pacific Oyster *Crassostrea gigas*. *G3 (Bethesda)* 10, 4691–4705. doi:10.1534/g3.120.401728

Yoshida, G. M., Bangera, R., Carvalheiro, R., Correa, K., Figueroa, R., Lhorente, J. P., et al. (2018). Genomic Prediction Accuracy for Resistance against *Piscirickettsia salmonis* in Farmed Rainbow Trout. *G3 (Bethesda)* 8, 719–726. doi:10.1534/g3.117.300499

Yoshida, G. M., Lhorente, J. P., Correa, K., Soto, J., Salas, D., and Yáñez, J. M. (2019). Genome-Wide Association Study and Cost-Efficient Genomic Predictions for Growth and Fillet Yield in Nile Tilapia (*Oreochromis niloticus*). *G3 (Bethesda)* 9, 2597–2607. doi:10.1534/g3.119.400116

Yuehuan, Z., Wu, X., Qin, Y., Xiao, S., Ma, H., Li, J., et al. (2017). Sustained Response to Selection of Growth Traits to the Third Generation for Two Strains of Kumamoto Oyster *Crassostrea sikamea*. *J. Fish. Sci. China* 24, 1161–1167. doi:10.3724/SP.J.1118.2017.16350

Zenger, K. R., Khatkar, M. S., Jones, D. B., Khalilisamani, N., Jerry, D. R., and Raadsma, H. W. (2019). Genomic Selection in Aquaculture: Application, Limitations and Opportunities with Special Reference to Marine Shrimp and Pearl Oysters. *Front. Genet.* 9, 693. doi:10.3389/fgene.2018.00693

Zhang, J., Li, Q., Xu, C., and Han, Z. (2019). Response to Selection for Growth in Three Selected Strains of the Pacific Oyster *Crassostrea gigas*. *Aquaculture* 503, 34–39. doi:10.1016/j.aquaculture.2018.12.076

Zhou, X., and Stephens, M. (2012). Genome-wide Efficient Mixed-Model Analysis for Association Studies. *Nat. Genet.* 44, 821–824. doi:10.1038/ng.2310

**Conflict of Interest:** Author RH was employed by Benchmark Genetics. Author JH was employed by Lochnell oysters.

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

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 18 July 2022
DOI 10.3389/fgene.2022.938473



OPEN ACCESS

EDITED BY
Md Samsul Alam,
Bangladesh Agricultural University, Bangladesh

REVIEWED BY
Poongodi Geetha-Loganathan,
SUNY Oswego, United States
Almas Gheyas,
University of Stirling, United Kingdom

*CORRESPONDENCE
Yongshuang Xiao,
dahaishuang1982@163.com
Jun Li,
junli@qdio.ac.cn,
lijun196412@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

RECEIVED 07 May 2022
ACCEPTED 28 June 2022
PUBLISHED 18 July 2022

CITATION
Ma Y, Xiao Y, Xiao Z, Wu Y, Zhao H, Gao G, Wu L, Wang T, Zhao N and Li J (2022), Genome-wide identification, characterization and expression analysis of the BMP family associated with beak-like teeth in Oplegnathus. *Front. Genet.* 13:938473. doi: 10.3389/fgene.2022.938473



# Genome-wide identification, characterization and expression analysis of the BMP family associated with beak-like teeth in *Oplegnathus*

Yuting Ma[1,2†], Yongshuang Xiao[2,3,4*†], Zhizhong Xiao[2,3,4,5,6], Yanduo Wu[2,5], Haixia Zhao[2,5], Guang Gao[2,5], Lele Wu[2,5], Tao Wang[1,2], Ning Zhao[2,5] and Jun Li[2,3,4*]

[1]School of Marine Science and Engineering, Qingdao Agricultural University, Qingdao, China, [2]CAS and Shandong Province Key Laboratory of Experimental Marine Biology, Center for Ocean Mega-Science, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China, [3]Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China, [4]Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, [5]College of Marine Science, University of Chinese Academy of Sciences, Beijing, China, [6]Weihai Haohuigan Marine Biotechnology Co., Weihai, China

Bone morphogenetic proteins (BMPs), which belong to the transforming growth factor beta (TGF-β) family, are critical for the control of developmental processes such as dorsal-ventral axis formation, somite and tooth formation, skeletal development, and limb formation. Despite *Oplegnathus* having typical healing beak-like teeth and tooth development showing a trend from discrete to healing, the potential role of BMPs in the development of the beak-like teeth is incompletely understood. In the present study, 19 and 16 BMP genes were found in *O. fasciatus* and *O. punctatus*, respectively, and divided into the BMP2/4/16, BMP5/6/7/8, BMP9/10, BMP12/13/14, BMP3/15 and BMP11 subfamilies. Similar TGFb and TGF_β gene domains and conserved protein motifs were found in the same subfamily; furthermore, two common tandem repeat genes (BMP9 and BMP3a-1) were identified in both *Oplegnathus fasciatus* and *Oplegnathus punctatus*. Selection pressure analysis revealed 13 amino acid sites in the transmembrane region of BMP3, BMP7, and BMP9 proteins of *O. fasciatus* and *O. punctatus*, which may be related to the diversity and functional differentiation of genes within the BMP family. The qPCR-based developmental/temporal expression patterns of BMPs showed a trend of high expression at 30 days past hatching (dph), which exactly corresponds to the ossification period of the bones and beak-like teeth in *Oplegnathus*. Tissue-specific expression was found for the BMP4 gene, which was upregulated in the epithelial and mesenchymal tissues of the beak-like teeth, suggesting that it also plays a regulatory role in the development of the beak-like teeth in *O. punctatus*. Our investigation not only provides a scientific basis for comprehensively understanding the BMP gene family but also helps screen the key genes responsible for beak-like tooth healing in *O. punctatus* and sheds light on the developmental regulatory mechanism.

KEYWORDS

*Oplegnathus fasciatus*, *Oplegnathus punctatus*, tooth development, BMP, gene family, adaptive evolution, gene expression

## Introduction

Bone morphogenetic proteins (BMPs) are potent growth factors forming the largest subfamily of the transforming growth factor beta (TGF-β) superfamily. The first BMP was isolated from demineralized bone tissue and named for its ability to induce ectopic endochondral osteogenesis after transplantation into rodent soft tissue (Wozney et al., 1988). To date, over 30 members have been identified in humans, with varying functions during processes such as embryogenesis, skeletal formation, haematopoiesis and neurogenesis (Ducy and Karsenty, 2000; Bragdon et al., 2011). Based on the structural homology and known functions of BMPs, they can typically be classified into five subgroups: the BMP2/4, BMP5/6/7/8, BMP9/10, and BMP12/13/14 groups (von Bubnoff and Cho, 2001; Mazerbourg and Hsueh, 2006). Unlike other BMP family members, BMP1 has a unique protein domain and is not a TGF-β superfamily member but a shrimp hubstopene family member (Bond and Beynon, 1995; Xie et al., 2020). However, studies in zebrafish have shown the importance of BMP1 for bone formation and stability (Asharani et al., 2012). The expression patterns of the BMP gene family and its roles in the bone metabolism and skeletal development of vertebrates have been reported. BMP-2 functions to regulate early dorsal-ventral patterning in vertebrate embryonic development (Rafael et al., 2006). BMP2/4/16 family members participate in various physiological processes during early embryonic development and are highly expressed in scale tissue. However, BMP16 is present in teleost fish but not in other tetrapods (Marques et al., 2016). BMP4 is secreted and synthesized by osteoblasts and can induce the differentiation and proliferation of undifferentiated mesenchymal cells and the formation of cartilage and new bone (Cheng et al., 2003; Canalis, 2009). BMP3 inhibits osteocyte differentiation and can negatively modulate bone density (Daluiski et al., 2001). Another report showed that BMP3 promotes the proliferation of mesenchymal stem cells through the TGF-β signaling pathway (Zhou et al., 2015) BMP9 has been reported to be involved in liver physiology and pathology (Desroches-Castan et al., 2019). For example, it promotes the proliferation, survival, invasion and cancer stem cell properties of hepatocellular carcinoma (HCC) cells (Li et al., 2013). BMP10 is essential for maintaining cardiac growth during murine cardiogenesis (Chen et al., 2004). BMP15 is an oocyte-specific growth factor that is specifically essential for female fertility. It co-regulates folliculogenesis and the ovulation rate with GDF9 (Galloway et al., 2000; Patino et al., 2017). BMP11 (also called growth differentiation factor 11, GDF11) is closely related to GDF-8 (myostatin) (Geng et al., 2010). BMP11 plays a critical role in regulating the axial skeleton in murine bone morphogenesis as well as promoting the formation of mesoderm and neural tissue (Gamer et al., 1999; McPherron et al., 1999). BMP-14 (also known as growth differentiation factor 5, GDF5) plays a crucial role in inducing the formation of ectopic cartilage (Hotten et al., 1996). Additional studies have shown that GDF5 can promote bone formation, cartilage formation, and longitudinal bone growth in the extremities (Tsumaki et al., 1999). BMP14 (also known as GDF5), BMP13 (also known as GDF6) and BMP12 (also known as GDF7) play important roles in the repair and regeneration of tendon/ligament injuries in rats (Wolfman et al., 1997). Studies have shown that BMP5 induces cartilage formation, and soft tissue defects in mice arise from a lack of BMP5 signaling in the subepithelial mesenchyme (King et al., 1994). BMP6 and BMP7 are involved in the formation and compartmentalization of the mid-heart cushion during mouse development (Kim et al., 2001). Moreover, BMP6 and BMP7 also play important roles in kidney diseases (Dendooven et al., 2011; Goncalves and Zeller, 2011). Studies have shown that BMP-8 may be an important player in bone metabolism, particularly in the response to glucocorticoids (Kosa et al., 2011).

The striped knifejaw, *Oplegnathus fasciatus*, and the spotted knifejaw, *Oplegnathus punctatus*, the latter of which is well known for its high nutritional and economic value, are economically important marine fishes in China (Bai et al., 2021). Due to its economic importance, *Oplegnathus* has become the subject of extensive research in various fields such as ecology, physiology, nutrition and evolutionary genomics (Xiao et al., 2020; Gong et al., 2022a; Gong et al., 2022b). *O. fasciatus* and *O. punctatus* inhabit rocky and coral reef areas offshore. They are carnivorous fishes with sharp teeth that allow them to bite through hard shells such as those of shellfish or sea urchins. Both *O. fasciatus* and *O. punctatus* have a distinctive parrot-like beak formed by the fusion of small teeth and tooth germs on the lower jaw (Kakizawa et al., 1980). All species of the *Oplegnathus* family have these highly specialized beak-like teeth, which help them catch and chew hard-bodied prey, such as decapods, echinoderms, and molluscs. *Oplegnathus woodwardi* consumes a substantial volume of sponges, which is facilitated by its possession of beak-like teeth (Maschette et al., 2020). The teeth of *Oplegnathus* heal with the upper and lower jaws to form strong beak-like teeth. Because of their strong teeth and biting ability, *Oplegnathus* fishes have also been called the "king of iso fishing" and are among the most challenging and interesting fishes to catch. Members of the teleost order *Tetraodontiformes*, such as *Triodontidae* (three-toothed pufferfishes), *Molidae* (ocean sunfishes), *Diodontidae* (porcupinefishes), and *Tetraodontidae* (four-toothed pufferfishes), have also evolved a variety of morphologically distinct beak-like teeth (Andreucci et al., 1982), and these unique and diverse teeth facilitate wide dietary niche occupancy, including access to hard-shelled prey (Turingan, 1994). BMPs, especially BMP2, BMP4, and BMP7, are essential for tooth development, not only inducing the proliferation of mesenchymal cells in the initial process of tooth germ

development and participating in the construction of developmental signalling centres but also functioning as a basic protein for the terminal differentiation of ameloblasts and odontoblasts (Liu and Lian, 2018; Liu and Lian, 2018). To further explore the role of BMPs in the development of beak-like teeth, genome-wide identification and annotation of BMP genes in *Oplegnathus* were carried out in this study. In addition, the genetic structure and evolutionary characteristics of these genes as well as their expression characteristics in different tissues and different developmental stages were also analysed. First, from an ecological point of view, the BMP gene family is the key regulator of beak-like tooth development, and the BMP/Smads pathway is involved in the regulation of osteoblast differentiation and bone formation. For example, the BMP2, BMP4 and BMP7 genes are essential for tooth development. Moreover, Oplegnathidae fishes have special healing beak-like teeth, so we were interested in investigating the regulatory roles of BMPs in beak-like tooth development and the healing process. Second, at the evolutionary level, we selected several species of fish with and without healing teeth and some representative higher animals as references to investigate potential differences in BMP family composition through evolutionary analysis.

# Materials and methods

## Ethics statement

All experiments strictly abided by the experimental animal guidance policy of the experimental animal ethics committee of the Institute of Oceanology, Chinese Academy of Sciences (Permit Number: IOCAS2022012).

## Experimental materials

*O. punctatus* were obtained from the Fish Breeding Base of The Institute of Oceanology, Chinese Academy of Sciences (Haihe Aquatic Seedling Co., Ltd., Weihai Wendeng). To detect the differential expression of BMP genes among the tissues of *O. punctatus*, we anaesthetized three adult *O. punctatus* with MS-222, dissected and collected eight different tissues (gills, liver, spleen, brain, the surrounding tissue of the beak-like teeth, intestine, muscle, and heart), and preserved the tissue samples in liquid nitrogen for further analysis.

## Morphological observation of the beak-like teeth of *Oplegnathus*

Three-year-old *Oplegnathus* samples were selected, their teeth were dissected, and the surface structure of the beak-like teeth was observed under a ZEISS Stemi 2000-c stereoscope. Otherwise, samples were imaged with high-resolution, high-contrast scans by a ZEISS Xradia 515 Versa 3D X-ray microscope at 20.0 and 3.0 μm voxel resolutions. Afterwards, ZEISS 3D Viewer software was used to observe the reconstructed 3D structure of the beak-like teeth of *Oplegnathus*, and the teeth were rendered in 3D using Dragonfly software. Samples of *O. punctatus* juveniles were collected for skeletal staining observation. Bone staining was performed according to the methods described by Dingerkus and Uhler (1977). Samples were stored in glycerol after processing, and a small amount of thymol was added to prevent bacterial contamination. Bone-stained *O. punctatus* samples were placed in a Petri dish with glycerol, observed and photographed with a ZEISS Stemi 2000-c stereoscope.

## Identification and sequence analysis of bone morphogenetic protein genes

All available BMP gene sequences and BMP amino acid sequences of nine species (*Cyprinus carpio*, *Salmo salar*, *Danio rerio*, *Oryzias latipes*, *Xenopus laevis*, *Takifugu rubripes*, *Gallus gallus*, *Mus musculus*, and *Homo sapiens*) were downloaded from the NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and Ensembl (http://asia.ensembl.org/) public databases. A simple HMM search in TBtools software and NCBI Blast were used to search the reference genomes of *O. fasciatus* and *O. punctatus* and preliminarily obtain candidate BMP gene sequences of these two species (Chen et al., 2020). Meanwhile, the NCBI-CDD database (https://www.ncbi.nlm.nih.gov/cdd/) was used to verify the identified characteristic domains of BMP genes in *O. fasciatus* and *O. punctatus*. The SMART (http://smart.embl-heidelberg.de) website was used to predict the structural information of BMP protein domains, such as TGFb and TGF_β. The mass fractions and isoelectric points of the BMP proteins were estimated by the ExPASy online tool (https://web.expasy.org/compute_pi). The reference genome accession codes used for analysis in this study are as follows: *O. fasciatus* (NCBI accession code: PRJNA393383) and *O. punctatus* (CNGB accession code: CNP0001488) (Xiao et al., 2019; Li et al., 2021).

## Phylogenetic analysis and chromosomal locations of bone morphogenetic protein genes

The BMP protein sequences of *O. fasciatus*, *O. punctatus* and the other nine species (*C. carpio*, *S. salar*, *D. rerio*, *O. latipes*, *X. laevis*, *T. rubripes*, *G. gallus*, *M. musculus*, and *H. sapiens*) were aligned using MUSCLE of TBtools (Edgar, 2004). Then, the created alignment file was used for phylogenetic analysis with MEGA11 software to construct a neighbour-joining (NJ) tree and a maximum likelihood (ML) tree (1,000 bootstraps) for comparison (Tamura et al., 2021). The online software EvolView (https://evolgenius.info//evolview-v2/) was used to beautify the evolutionary tree. Multiple sequence alignment of *O. fasciatus* was performed using Jalview software. MEME (http://meme-suite.org/tools/meme) online

software was used to analyse BMP motifs. Then, the results were visualized with TBtools software. The conserved domains of BMPs in *Oplegnathus* were obtained using CDD of the NCBI and visualized by TBtools. Chromosome position and gene density information were obtained through the *Oplegnathus* genome database, the BMP gene positions were marked on the chromosomes, the collinear relationships between different chromosomes were obtained through the genome database, and the results were visualized using TBtools software.

## Evolutionary selection pressure analysis of bone morphogenetic protein genes

Phylogenetic trees were constructed individually for BMP gene families from 11 species using the NJ method in MEGA11. Using the CODEML program of PAML4.9j (Yang, 2007), selection pressure analysis was performed on the main clades using different models (branch model, site model and branch-site model). First, branch models are used to determine whether the selection pressure on each branch of the evolutionary tree is significant (Yang, 1998). The main branch models are the one-ratio model (M0), in which the ω-values of all branches in the phylogenetic tree are equal; the free-ratio model (M1), in which the ω-values among branches in the phylogenetic tree are unequal; and the two-ratio model (M2), where the ω-values of foreground branches are different from those of background branches. In addition, the site model assumes that different branches of the phylogenetic tree are subject to the same selection pressure but that different amino acid sites experience different selection pressures. The following are the main site models: the one ratio model (M0), assuming that all sites have the same ω value; the near-neutral model (M1a), assuming that only conserved sites ($0 < \omega < 1$) and neutral sites ($\omega = 1$) are present; the positive selection model (M2a), in which sites with positive selection ($\omega > 1$) are added to M1a; the discrete model (M3), assuming a simple discrete distribution trend of ω values for all sites; the beta model (M7), assuming that $\omega$ belongs to the matrix (0, 1) and follows a beta distribution for all sites; and the beta and ω model (M8), in which another class of ω values is added to M7, which can be obtained computationally and can be set to 1. Comparisons between M1 and M2, M0 and M3, and M7 and M8 are often used for site models, and M1 and M2 are more stable than M7 and M8. Moreover, in the analysis of the branch-site model, we tested for positively selected amino acid sites among the branches through the settings of different foreground branches and compared the null hypothesis of each Model A (MA) with the corresponding alternative hypothesis (Zhang et al., 2005). Then, the ratio ω of non-synonymous to synonymous substitutions (dN/dS) was calculated. When $\omega = 1$, selection is neutral (that is, there is no selection); when $\omega > 1$, there is positive selection; and when $0 < \omega < 1$, there is negative selection, also called purifying selection. In this paper, the likelihood ratio test (LRT) was used to compare model pairs, and the Bayesian method (BEB) was used to identify the sites with positive selection (Bielawski and Yang, 2003).

## Expression of bone morphogenetic protein genes

RT–PCR was performed with the CFX96 real-time detection system (Bio-Rad, United States) to determine BMP expression differences in different developmental stages and tissues of *O. punctatus*. Total RNA was isolated from samples of eight different tissues (gill, liver, spleen, brain, the surrounding tissue of the beak-like teeth, intestine, muscle, and heart) and five developmental stages (5 days past hatching (dph), 16, 30, 50, and 60 dph) in *O. punctatus* using TRIzol (Invitrogen) according to the manufacturer's instructions. Considering that the fish were too small to precisely separate the surrounding tissue of the beak-like teeth, the sampling locations for the five developmental stages were on the head, except for days 5 and 16 dph, when whole fish were taken because they were too small to precisely separate the head. RNA concentration and purity were measured using an ultra-micro spectrophotometer (NanoDrop 2000, United States). In general, the OD260 nm ratio was OD280 nm ≥ 1.8, and the OD260 nm ratio was OD230 nm ≥ 1. RT-PCR was conducted to synthesize cDNA from 1 μg of total RNA by using an Evo M-MLV RT Kit (Accurate Biology, Changsha, China) following the manufacturer's instructions. The internal positive control was the β-actin gene from *O. punctatus*, with the forward primer (5′-GCTGTGCTGTCCCTGTA-3′) and reverse primer (5′-GAGTAGCCACGCTCTGTC-3′). Details of the BMP gene primer pairs are provided in Supplementary Table S1. qRT–PCR experiments were performed with a 20 μl mixture consisting of 0.4 μl of each primer, 10 μl of SYBR® Premix Ex Taq™ II (Tli RNaseH Plus) (2X), 7.2 μl of $ddH_2O$, and 2 μl of template cDNA. The PCR program consisted of an initial denaturation step at 95°C for 30 s and 39 denaturation cycles: denaturation at 95°C for 5 s and annealing at 60°C for 30 s. The temperature depended on the primers. Melting curve analysis was performed at the end of the reaction to demonstrate its specificity. Each experiment was performed in triplicate. β-Actin was used for normalization of all gene expression data (Wang et al., 2020). According to the measured Ct value (the average value of three parallel samples for Ct), the relative expression levels of the BMP genes were calculated by Pfaffl's method (Pfaffl, 2001). All data obtained in the experiment were analysed by one-way analysis of variance using GraphPad Prism 9 software, and $p < 0.05$ was indicative of a statistically significant difference. Quantitative results were plotted using the R environment and GraphPad Prism 9 software.

# Results

## Morphological observation of *Oplegnathus*

As shown in Figure 1, the *Oplegnathus* species have a unique healing beak-like teeth structure, with the upper and lower jaw

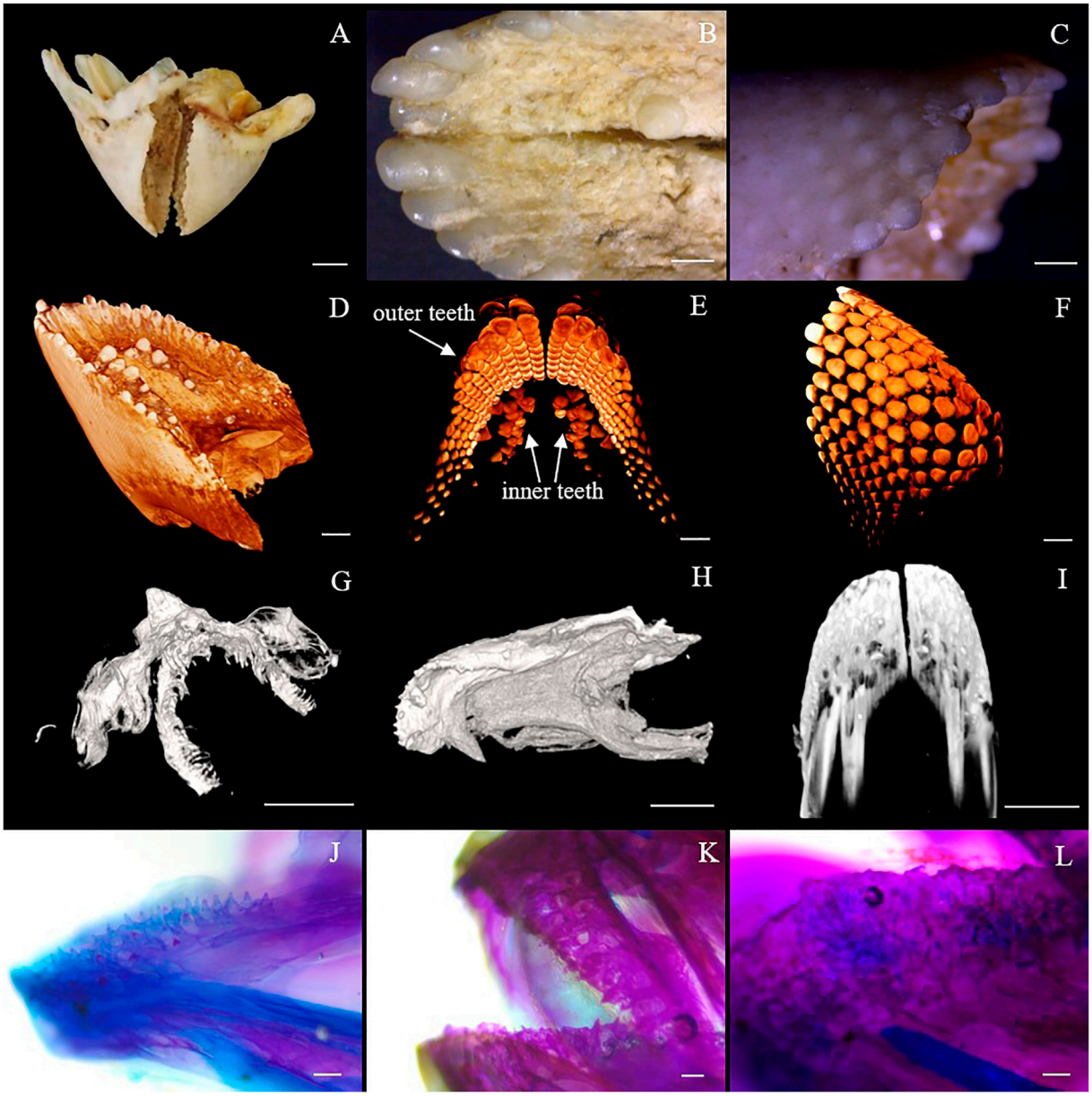



FIGURE 1
Beak-like tooth morphology of *Oplegnathus*. **(A)** Photographs of the overall morphology of the beak-like teeth of *Oplegnathus*. **(B)** Enlarged image of locoregional details of the inner teeth of *Oplegnathus*. **(C)** Enlarged image of locoregional details of the outer teeth of *Oplegnathus*. **(D–F)** Reconstructed micro-CT scans showing the nested dentition of adult *Oplegnathus*. **(E)** Three-dimensional arrangement of the inner and outer teeth of *Oplegnathus*. **(G–I)** Micro-CT scans showing the developmental changes of the nested dentition of *Oplegnathus* at 30, 50 and 70 dph. **(J–L)** Alizarin red staining results of beak-like teeth in different developmental stages of *Oplegnathus*. **(J)** indicates that the 45-dph beak-like teeth are in a cartilaginous state and separately arranged, **(K)** indicates that the 50-dph beak-like teeth are ossified and have a tendency to heal, and **(L)** indicates that complete healing of the beak-like teeth has occurred by 90 dph (scale bars: 5 mm in **(A)**; 1 mm in **(B,C,I)**; 2 mm in **(D,F)**; 500 μm in **(G,H)**; 100 μm in F and H; and 100 μm in **(J–L)**.

teeth showing healing and a gap in the middle. In addition, there are obvious traces of the accumulation of replacement teeth on the outer edges of the beak-like teeth. Moreover, the outer layer of the beak-like teeth is filled with smooth calcium, while the inner layer is relatively rough. As shown in Figure 1, the 3D rendering analysis of the thickness of the teeth by micro-CT revealed a nested internal arrangement of the teeth of the *Oplegnathus* species, extending from far from the rostral end to near the rostral end. In addition, several irregularly distributed internal teeth could be observed on the inner edge of the beak-like teeth (Figure 1E). The micro-CT scan results at 30, 50 and 70 dph showed beak-like teeth distributed in a single row of separation at

TABLE 1 Summary of the characteristics of BMP genes in *O. fasciatus* and *O. punctatus*.

| Gene name | Gene ID | Location | CDS (na) | CDS (aa) | PI | MW (Da) | Accession number |
|---|---|---|---|---|---|---|---|
| OfBMP 2 | Opa012474 | Chr11 | 1608 | 535 | 6.93 | 60,430.29 | ON881245 |
| OfBMP 3a | Opa007971 | Chr19 | 1323 | 440 | 9.6 | 50,769.68 | ON881248 |
| OfBMP 3b-1 | Opa022377 | Chr18 | 1395 | 464 | 9.32 | 52,078.46 | ON881246 |
| OfBMP 3b-2 | Opa018624 | Chr12 | 1605 | 534 | 9.3 | 60,351.16 | ON881247 |
| OfBMP 4 | Opa019617 | Chr15 | 1212 | 403 | 7.74 | 46,269.59 | ON881249 |
| OfBMP 5 | Opa018570 | Chr12 | 1359 | 452 | 8.96 | 50,980.73 | ON881250 |
| OfBMP 6 | Opa020205 | Chr16 | 1137 | 378 | 8.73 | 42,817.76 | ON881251 |
| OfBMP 7a | Opa003248 | Chr8 | 1278 | 425 | 5.98 | 48,584.1 | ON881252 |
| OfBMP 7b | Opa005368 | Chr4 | 1284 | 427 | 6.53 | 48,861.37 | ON881253 |
| OfBMP 8 | Opa015260 | Chr21 | 1362 | 453 | 9.19 | 51,095.91 | ON881254 |
| OfBMP 9 | Opa022376 | Chr18 | 966 | 321 | 5.81 | 36,617.86 | ON881255 |
| OfBMP 10a | Opa021035 | Chr17 | 1395 | 464 | 6.13 | 52,883.03 | ON881256 |
| OfBMP 10 | Opa008866 | Chr19 | 1488 | 495 | 5.14 | 55,413.19 | ON881257 |
| OfBMP 11-1 | Opa003044 | Chr8 | 1170 | 389 | 6.67 | 44,283.68 | ON881258 |
| OfBMP 11-2 | Ofa014976 | Chr9 | 1131 | 376 | 5.6 | 42,776.95 | ON881259 |
| OfBMP 11-3 | Ofa023623 | Chr1 | 1080 | 359 | 6.27 | 40,682.57 | ON881260 |
| OfBMP 13 | Opa010039 | Chr3 | 1335 | 444 | 9.25 | 50,612.51 | ON881261 |
| OfBMP 15 | Opa016633 | Chr14 | 1353 | 450 | 9.56 | 51,083.92 | ON881262 |
| OfBMP 16 | Opa006329 | Chr7 | 1443 | 480 | 9.49 | 53,148.13 | ON881263 |
| OpBMP2 | mikado.FChr_17G371.1 | FChr_17 | 1269 | 422 | 8.59 | 47,680.8 | ON881264 |
| OpBMP3b-1 | mikado.FChr_22G218.1 | FChr_22 | 1368 | 455 | 9.39 | 51,047.4 | ON881265 |
| OpBMP3b-2 | mikado.FChr_15G610.1 | FChr_15 | 1587 | 528 | 9.15 | 59,774.4 | ON881266 |
| OpBMP4 | mikado.FChr_20G478.1 | FChr_20 | 1419 | 472 | 8.38 | 54,316.1 | ON881267 |
| OpBMP6 | mikado.FChr_21G268.1 | FChr_21 | 1287 | 428 | 6.99 | 48,845.4 | ON881268 |
| OpBMP7a | mikado.FChr_3G1132.1 | FChr_3 | 1278 | 425 | 5.98 | 48,578.1 | ON881269 |
| OpBMP7b | mikado.FChr_7G612.1 | FChr_7 | 1971 | 656 | 6.4 | 73,424.9 | ON881270 |
| OpBMP8 | mikado.FChr_13G995.1 | FChr_13 | 1422 | 473 | 9.22 | 53,168.2 | ON881271 |
| OpBMP9 | mikado.FChr_22G219.1 | FChr_22 | 1335 | 444 | 5.9 | 50,131.1 | ON881272 |
| OpBMP10 | mikado.FChr_6G239.1 | FChr_6 | 1491 | 496 | 5.09 | 55,597.3 | ON881273 |
| OpBMP11-1 | mikado.FChr_3G854.1 | FChr_3 | 1167 | 388 | 6.67 | 44,185.6 | ON881274 |
| OpBMP11-2 | mikado.FChr_14G143.1 | FChr_14 | 1131 | 376 | 5.6 | 42,777 | ON881275 |
| OpBMP11-3 | mikado.FChr_24G619.1 | FChr_24 | 1107 | 368 | 6.47 | 41,654.7 | ON881276 |
| OpBMP13 | mikado.FChr_13G951.1 | FChr_13 | 1239 | 412 | 9.16 | 47,141 | ON881277 |
| OpBMP15 | mikado.FChr_16G452.1 | FChr_16 | 1353 | 450 | 9.45 | 51,100 | ON881278 |
| OpBMP16 | mikado.FChr_4G590.1 | FChr_4 | 1557 | 518 | 9.72 | 57,520.1 | ON881279 |

30 dph, replacement teeth at 50 dph, beak-like teeth that generally tended to heal, and complete rostral tooth healing at 70 dph (Figures 1G–I). The results of bone staining are shown in Figures 1J–L. At 45 dph, the beak-like teeth are individually arranged conical teeth, and most of the teeth and bones are stained blue with bromophenol blue, showing a cartilaginous state; only the tips of the teeth, which are beginning to ossify, are red. After 50 dph, the discrete arrangement of conical teeth shows a tendency to heal, and all of the beak-like teeth are stained with alizarin red, indicating that ossification has been completed. By 90 dph, the beak-like teeth of *O. punctatus* had healed. External observation showed clear signs of replacement tooth accumulation on the outer edges of the upper and lower teeth, with three to four rows of replacement teeth.

## Bone morphogenetic protein gene family identification

The cDNA sequence characteristics and accession numbers of all cloned BMP genes of *Oplegnathus* are shown in Table 1. The accession numbers of all cloned BMP genes of the other 11 species are shown in Supplementary Table S2. 19 and 16 different BMP genes were found in the genomes of *O.*

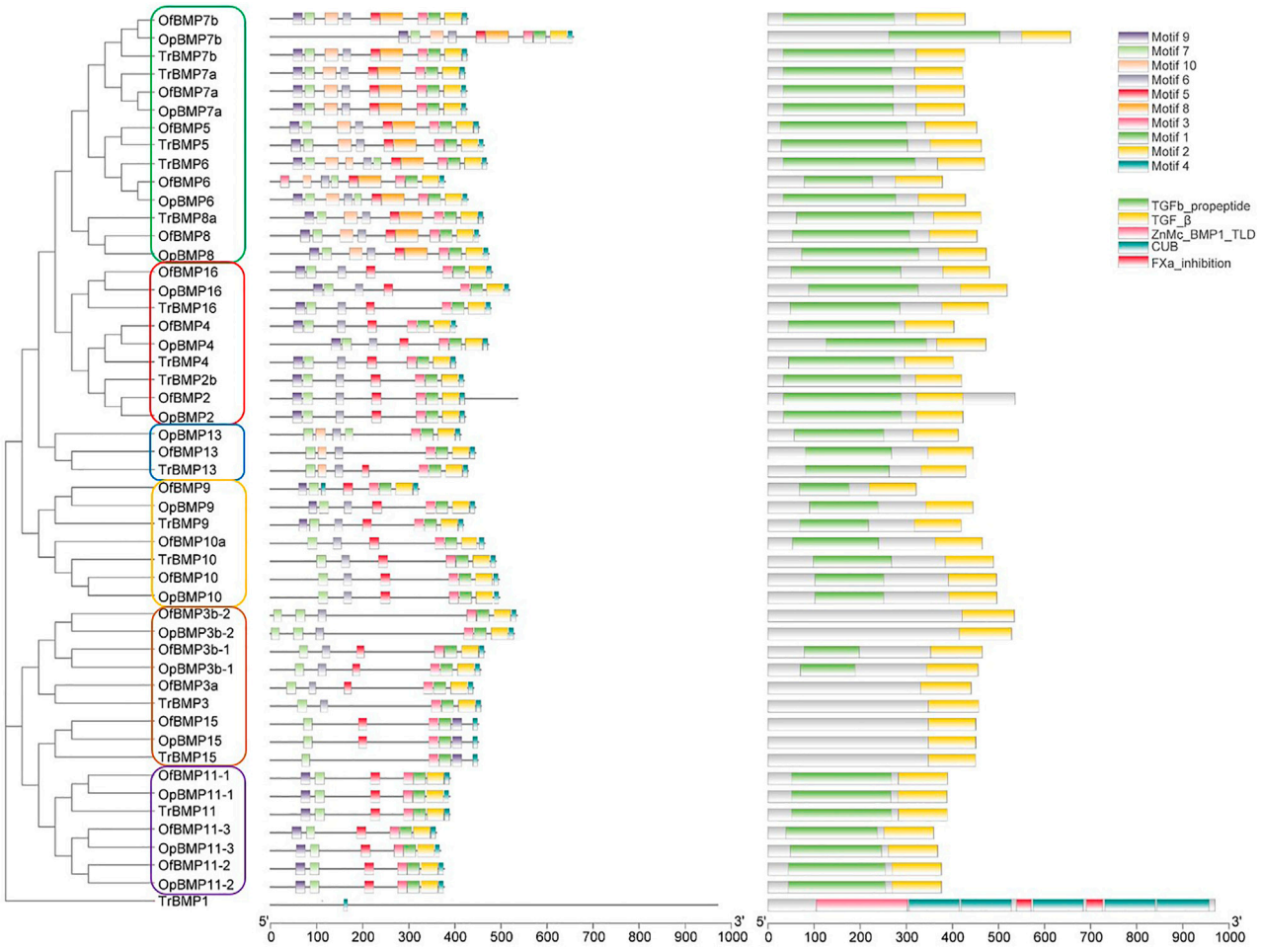



FIGURE 2
The conserved BMP domains and motifs in *Oplegnathus* and *T. rubripes*.

*fasciatus* and *O. punctatus*, respectively (Table 1). In Table 1, genes with the prefix "Of" before "BMP" are BMP genes in *O. fasciatus*, and those with the prefix "Op" are BMP genes in *O. punctatus*. The length of the coding sequence (CDS) is 966–1,971 bp, the length of the encoded protein is 321–656 amino acids, the molecular weight is between 36 and 73 kDa, and the isoelectric point is between 5.09 and 9.72. As shown in Figure 2, to deeply investigate the structural diversity of the BMP gene family, the conserved domains and motifs of the *Oplegnathus* BMP proteins were visualized by TBtools software. Except BMP1, the *Oplegnathus* BMPs have similar TGF_β domains, indicating that the BMP gene family belonging to the TGF-β superfamily is highly conserved. In addition, most BMP genes, except BMP3 and BMP15, also possess a TGFb propeptide domain, revealing their potentially differentiated functions in biological processes. Furthermore, a total of 10 distinct conserved motifs were identified. The conserved domains and motifs were compared according to the phylogenetic relationships between *Oplegnathus* and *T. rubripes* (please refer to Supplementary Figure S1 for a description of the names and lengths of the motifs). The highly consistent conserved domains and motifs in the same subfamilies indicated similar protein functions among members of the same subfamily. To explore the conserved domains and phylogenetic relationships of the BMP proteins of *O. fasciatus* and *O. punctatus*, multiple sequence alignment of their TGF_β domains was performed, revealing that all BMP family members contain a highly conserved TGF_β region at the C-terminus, which consists of approximately 100–110 amino acids (Figure 3). As shown in Figures 4A, 19 BMPs were distributed among 15 of the 24 chromosomes in *O. fasciatus*. Both BMP3b-1 and BMP9 were located on chromosome 10 and had tandem duplications. As shown in Figures 4B, 16 BMPs were distributed among 13 of the 24 chromosomes in *O. fasciatus*. Both BMP3b-1 and BMP9 were located on chromosome 22. As shown in Figure 5A, the homologous fragment of the OfBMP3b-

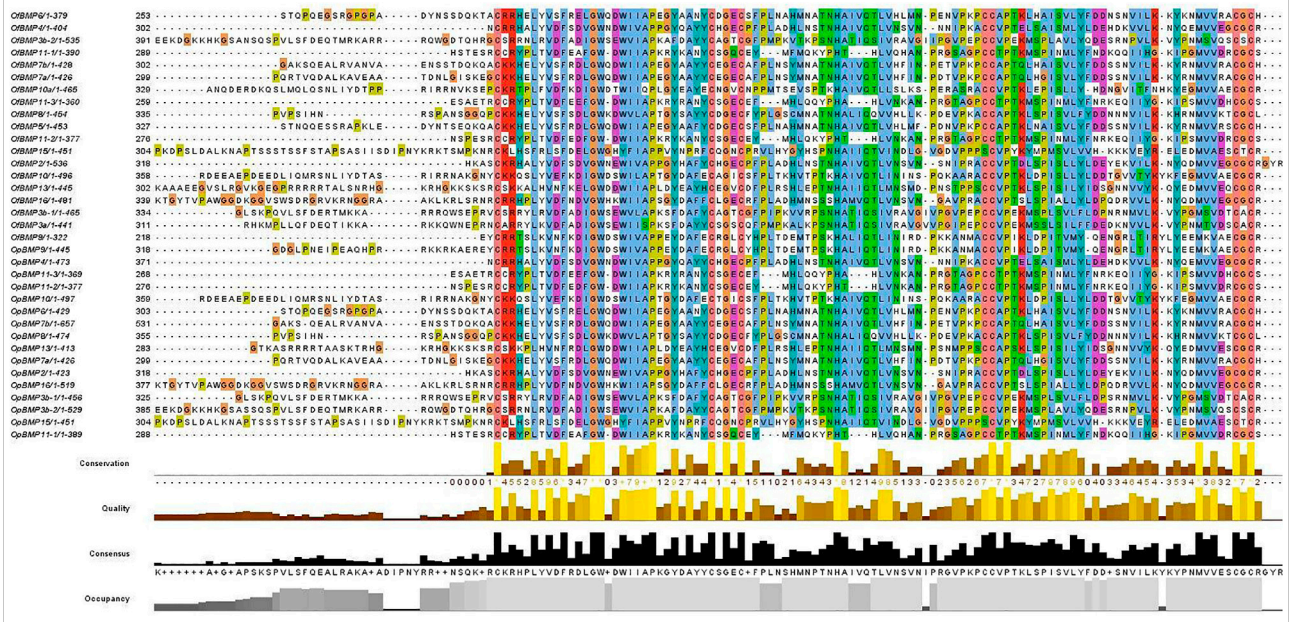



FIGURE 3
Multiple sequence alignment of the conserved domains of proteins in the *Oplegnathus* BMP gene family.

1 gene located on chromosome 18 of *O. fasciatus* is the OfBMP3b-2 gene located on chromosome 12. Moreover, the homologous fragment of the OfBMP7a gene located on chromosome 8 of *O. fasciatus* is the OfBMP7b gene located on chromosome 4. Furthermore, the homologous fragment of OfBMP13 on chromosome 3 of *O. fasciatus* is located on chromosome 21. As observed in *O. fasciatus*, the homologous fragment of the OpBMP3b-1 gene located on chromosome 22 in *O. punctatus* is the OpBMP3b-2 gene located on chromosome 15. Additionally, the homologous fragment of the OpBMP7a gene located on chromosome 3 in *O. punctatus* is the OpBMP7b gene located on chromosome 7. Furthermore, the homologous fragment of OpBMP13 on chromosome 13 of *O. punctatus* is located on chromosome 9 (Figure 5B).

## Phylogenetic analysis

To demonstrate the phylogenetic relationships of BMP genes between *Oplegnathus* and other species, we selected a total of 198 BMP amino acid sequences from 11 species, namely, *O. fasciatus*, *O. punctatus, C. carpio, S. salar, D. rerio, O. latipes, X. laevis, T. rubripes, G. gallus, M. musculus,* and *H. sapiens*. The phylogenetic tree was constructed by the ML method using MEGA11 and TBtools. As shown in Figure 6, the high topological consistency of the phylogenetic relationships of the BMPs indicates that they are highly evolutionarily conserved. Based on structural homology, except for BMP1, the BMPs were divided into six subfamilies, namely, the BMP2/4/16, BMP5/6/7/8, BMP9/10, BMP12/13/14, BMP3/15, and BMP11 groups. Both *O. fasciatus* and *O. punctatus* contain three members of the BMP11 gene family: BMP11-1, BMP11-2, and BMP11-3. *O. fasciatus* has three members of the BMP3 gene family: OfBMP3a, OfBMP3b-1, and OfBMP3b-2. *O. punctatus* has two members of the BMP2 gene family: OpBMP3b-1 and OpBMP3b-2. The copy number of each BMP gene in nine vertebrates was summarized and analysed. The details of the numbers of BMP members among the 11 species are provided in Table 2. Except in *C. carpio*, the number of BMP family members in the vertebrates ranges from 14 to 22. *C. carpio* (Xu et al., 2014; Chen et al., 2017) experienced an additional whole-genome duplication event, giving it approximately double the number of paralogous genes in zebrafish. As shown in Figure 7, BMP1, BMP12, and BMP14 were not found in *Oplegnathus* or *T. rubripes*. Additionally, BMP16 has been found only in fish species. BMP13 and 14 are absent in *X. laevis,* while BMP8 and BMP12 are absent in *G. gallus*.

## Evolutionary selection pressure analysis of bone morphogenetic protein genes

In the branch-site model, we set the branch on which the target BMP genes of *O. fasciatus* and *O. punctatus* were located as

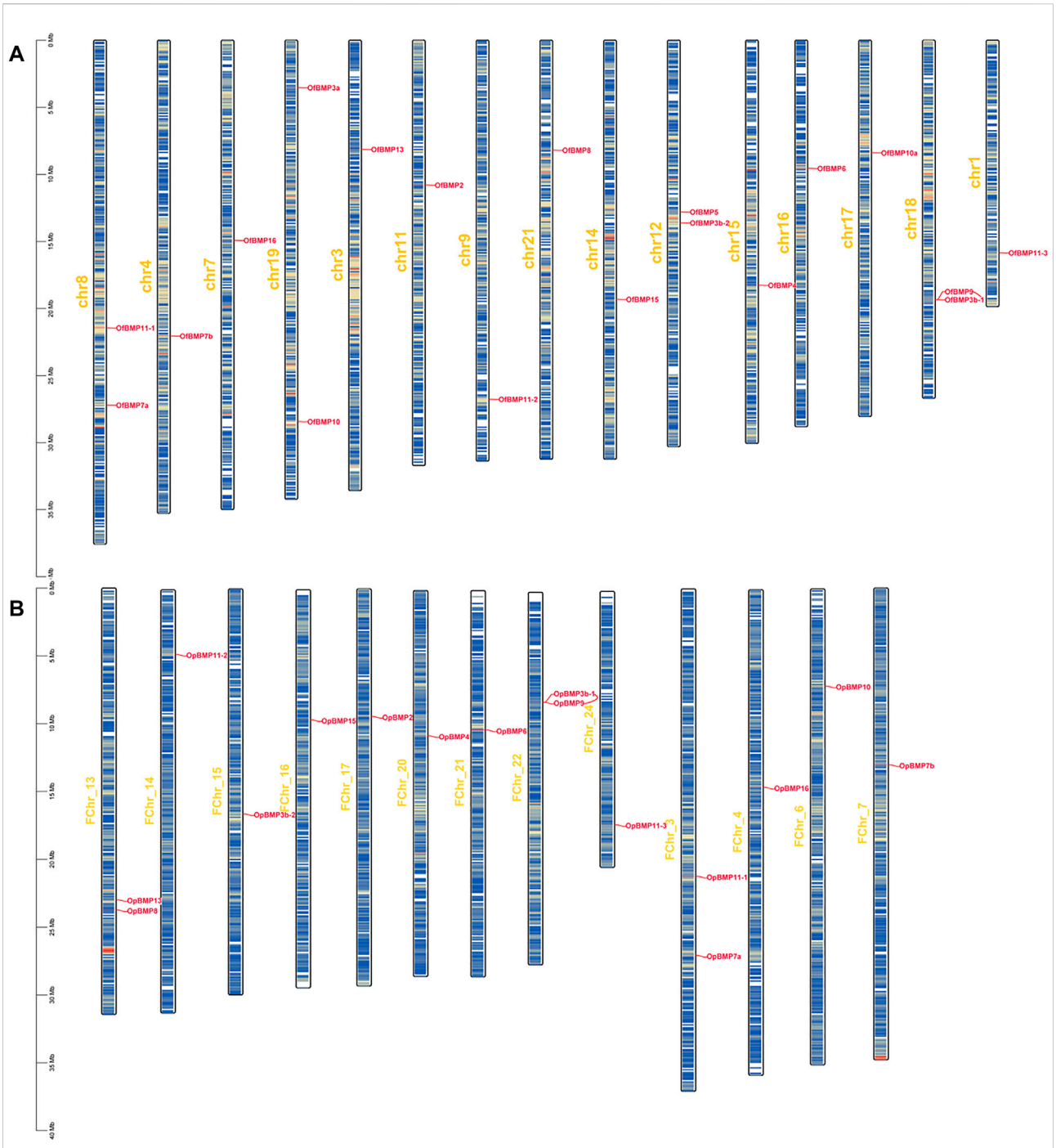



FIGURE 4
Chromosome localization of BMP genes in *O. fasciatus* and *O. punctatus*. **(A)** Represents *O. fasciatus*; **(B)** represents *O. punctatus*. The prefix "F" in the chromosome names in Panel **(B)** indicates that the reference genomic data are from female *O. punctatus*.

the foreground branch, and the branch on which the BMP genes of the other species were located was set as the background branch. According to the results in Table 3, we identified three genes (BMP3, BMP7, and BMP9) with positively selected sites under the branch-site model. In contrast, no positively selected sites were detected for BMPs in the branch model and site model, and the omega values were less than 1, indicating that the BMP genes are strongly negatively selected under this model. In the branch-site model, for the BMP3 gene, the following 11 amino acid sites were positively selected: 136H, 138K, 140V, 141F, 143F,

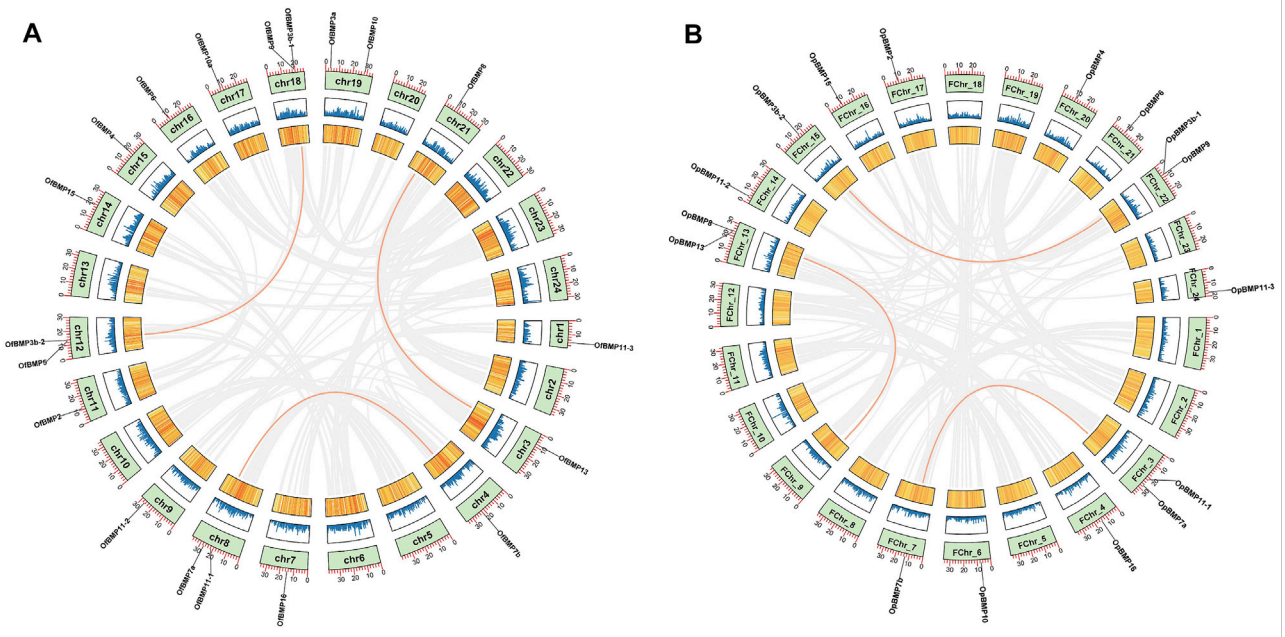



FIGURE 5
Chromosome localization and collinearity analysis of BMP genes in *O. fasciatus* and *O. punctatus*. **(A)** Represents *O. fasciatus*; **(B)** represents *O. punctatus*.

145L, 146S, 148I, 150E, 151S, and 153L. Similarly, 559T of BMP7 and 26F of BMP9 were identified as positively selected sites. Furthermore, the LRTs show that the above results are reliable, and the alternative hypothesis Model A is accepted. The specific parameters and results of the model are shown in Supplementary Tables S4–S9. Please refer to the Supplementary Material for details on the model parameters and LRTs for other BMP genes.

## Tissue-specific expression of bone morphogenetic protein genes in *Oplegnathus punctatus*

The qRT–PCR expression results showed that the BMP genes were expressed in various healthy tissues of *O. punctatus*, but the expression patterns of these genes varied among the tissues. The gill expression values were selected as a reference for comparison with the expression values of other tissues during the analysis of PCR gene expression data, so the relative expression value for gill tissue was set at 1. In general, most BMP genes were widely expressed, and there was a trend for higher relative expression of most BMP genes in the gill, as this tissue is more likely to be associated with BMP-related pathways. As shown in Figure 8, regarding our focus on the epithelial tissue and mesenchymal tissue of the beak-like teeth, the high relative expression of BMP2, BMP3b-1, BMP3b-2, BMP4, and BMP11-1 in the epithelial tissue and mesenchymal tissue of the beak-like teeth (all values above 1) is interesting. However, the relative expression of BMP3b-2 in the spleen and brain was higher than that in the epithelial tissue and mesenchymal tissue of the beak-like teeth. The relative expression of the remaining BMP genes in the epithelial tissue and mesenchymal tissue of the beak-like teeth was low (all below 0.6). Based on the tissue expression patterns of BMP genes, several BMP genes (BMP2, BMP3b-1, BMP3b-2, BMP4, and BMP11-1) that were highly expressed in the surrounding tissue of the beak-like teeth were screened. As shown in Figure 10, their developmental expression patterns at 30, 50, and 60 dph were also analyzed in correlation with the results of alizarin red staining at early developmental stages. As shown in Figures 8A–C, there are obvious differences in the expression levels of the BMP2/4/16 subfamily. The expression abundance of BMP2 was higher in the gills, liver, intestine and the epithelial tissue and mesenchymal tissue of the beak-like teeth, followed by the spleen and brain, with lower levels in the muscle and heart. BMP3b-1 and BMP4 exhibited similar expression patterns; however, their expression in the liver was significantly lower than that of BMP2. In contrast, BMP16 was highly expressed in the muscle, heart and gill, while the lowest expression was detected in the epithelial tissue and mesenchymal tissue of the beak-like teeth. Unlike BMP3b-1, BMP3b-2 is most highly expressed in the brain, followed by the spleen and gill, and weakly expressed in the muscle and heart. As shown in Figure 8B,

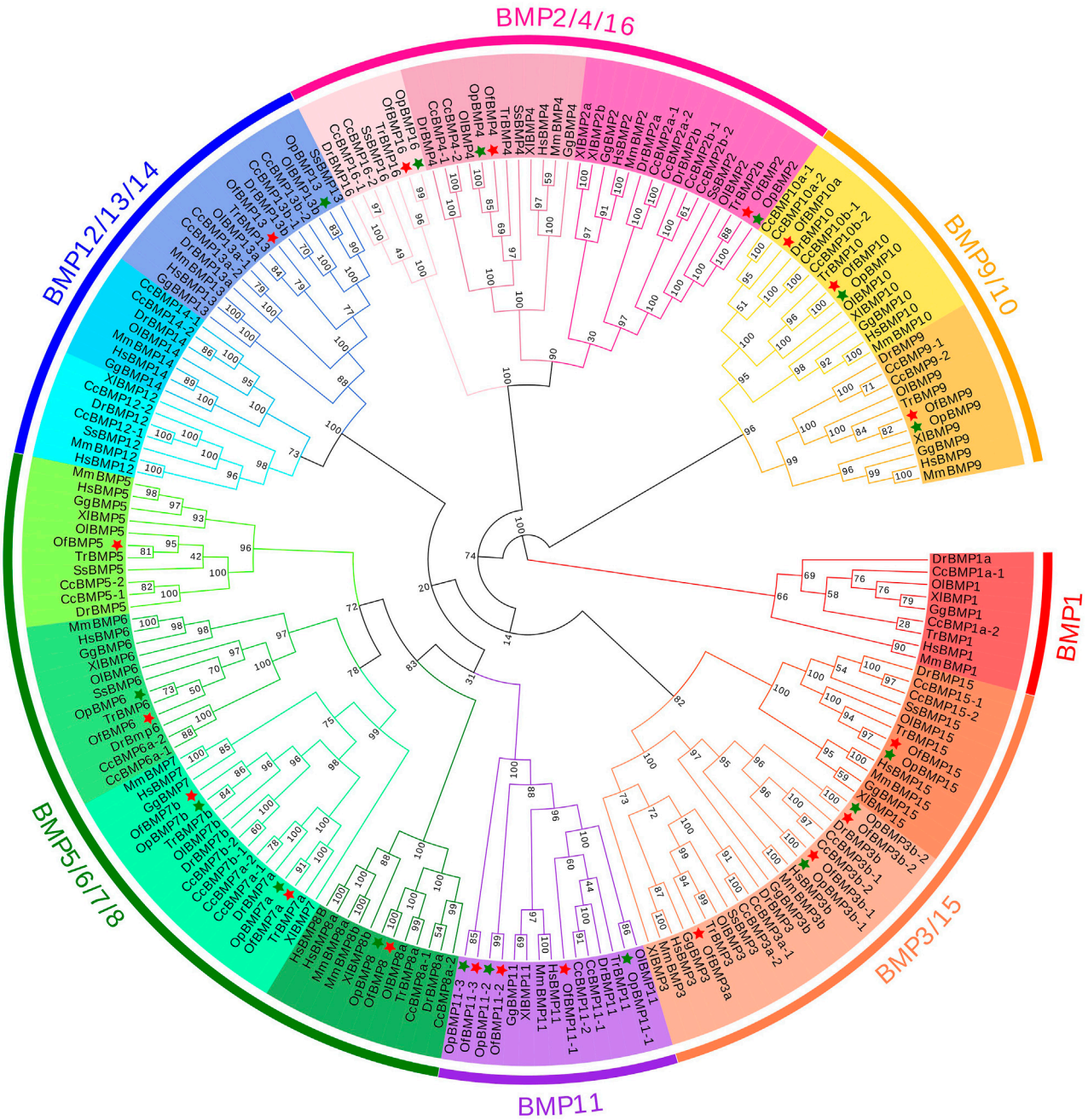



FIGURE 6
Phylogenetic tree of BMP gene family members in 11 species.

the BMP5/6/7/8 subfamily exhibited various expression levels across the examined tissues. There was significant downregulation in the epithelial tissue and mesenchymal tissue of the beak-like teeth for the BMP6, BMP7a, BMP7b, and BMP8 genes. The expression of BMP6 and BMP8 was the highest in the gill. BMP7a was highly expressed in the brain, while BMP7b was highly expressed in the muscle. As shown in Figure 8C, the BMP9/10 subfamily showed higher expression in the gill, muscle and heart. For BMP9, the expression level was the lowest in the liver, followed by the brain, spleen, the surrounding tissues of beak-like teeth, and intestine. The expression of BMP10 was higher in the liver, followed by the spleen, brain and intestine, and the lowest expression was observed in the surrounding tissues of the beak-like teeth. As shown in

TABLE 2 Comparison of copy numbers of BMP genes among selected vertebrate genomes.

| | *O. fasciatus* | *O. punctatus* | *C. carpio* | *S. salar* | *D. rerio* | *O. latipes* | *T. rubripes* | *X. laevis* | *G. gallus* | *M. musculus* | *H. sapiens* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| BMP1 | | | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| BMP2 | 1 | 1 | 4 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 |
| BMP3 | 3 | 2 | 4 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 2 |
| BMP4 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| BMP5 | 1 | | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| BMP6 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| BMP7 | 2 | 2 | 4 | | 2 | 1 | 2 | 1 | 1 | 1 | 1 |
| BMP8 | 1 | 1 | 2 | | 1 | 1 | 1 | 1 | | 2 | 2 |
| BMP9 | 1 | 1 | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| BMP10 | 2 | 1 | 4 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| BMP11 | 3 | 3 | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| BMP12 | | | 2 | 1 | 1 | | | 1 | | 1 | 1 |
| BMP13 | 1 | 1 | 4 | 1 | 2 | 2 | 1 | | 1 | 1 | 1 |
| BMP14 | | | 2 | | 1 | 1 | | | 1 | 1 | 1 |
| BMP15 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| BMP16 | 1 | 1 | 2 | 1 | 1 | | 1 | | | | |
| Total | 19 | 16 | 42 | 9 | 20 | 15 | 15 | 14 | 14 | 17 | 17 |

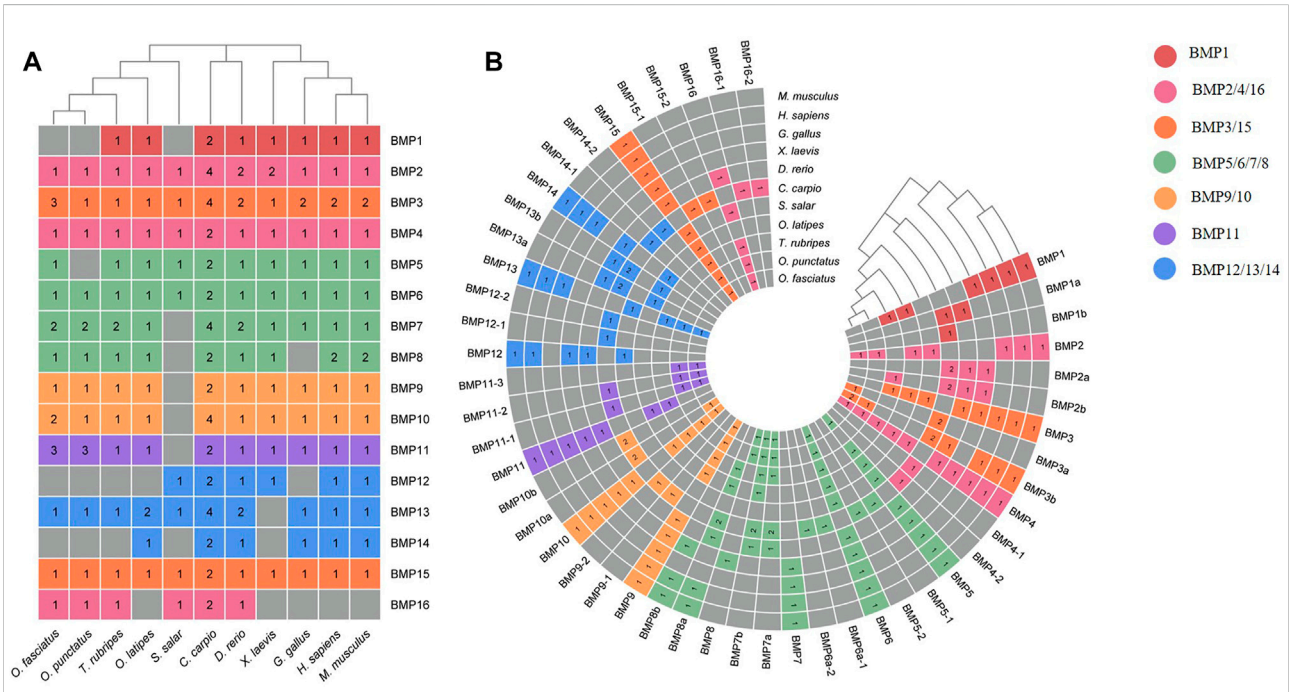



FIGURE 7
Heatmap of gene copy number comparisons of BMP genes in 11 species. **(A)** Shows a summary of the gene copy number comparison. **(B)** Shows the detailed gene copy number comparison. The different colors in the Figure refer to different BMP subfamilies, the numbers on the heatmap indicate the number of gene copies, and the gray color indicates the deletion of the BMP genes.

Figure 8C, members of the BMP11 subfamily were mainly expressed in the gill. Additionally, BMP11-1 was also highly expressed in the intestine and the surrounding tissues of the beak-like teeth, while BMP11-2 and BMP11-3 were highly expressed in the muscle and heart. Furthermore, there was significant downregulation in the liver for BMP11-2.

TABLE 3 Likelihood ratio test statistics for BMP3, BMP7 and BMP9.

| | Model | Model comparison | Positive selection site (BEB) | df | LRT (2ΔL) | *p* | Accepted model |
|---|---|---|---|---|---|---|---|
| | Branch model | M0 vs. M1 | | 36 | 97.986196 | 0.00000012 | M1 |
| | | M0 vs. M2-3a | | 1 | 6.545132 | 0.01051734 | M2 |
| | | M0 vs. M2-3b | | 1 | 0.012826 | 0.90992187 | M0 |
| | Site model | M0 vs. M3 | | 4 | 1113.924544 | 0 | M0 |
| | | M1a vs. M2a | | 2 | 0 | 1.00000000 | M1a |
| | | M7 vs. M8 | | 2 | 2.606358 | 0.27167467 | M7 |
| | | MA vs. null-3a | 136 H 0.978* | | | | |
| | | | 138 K 0.993** | | | | |
| BMP3 | | | 140 V 0.994** | | | | |
| | | | 141 F 0.987* | | | | |
| | | | 143 F 0.956* | | | | |
| | Branch-site model | | 145 L 0.971* | 1 | 56.971496 | 0.00000000 | MA |
| | | | 146 S 0.998** | | | | |
| | | | 148 I 0.996** | | | | |
| | | | 150 E 0.988* | | | | |
| | | | 151 S 0.998** | | | | |
| | | | 153 L 0.995** | | | | |
| | | MA vs. null-3b | | 1 | 7.507388 | 0.00616990 | MA |
| | Branch model | M0 vs. M1 | | 30 | 133.166076 | 0.00000000 | M1 |
| | | M0 vs. M2-7a | | 3 | 11.228174 | 0.01055428 | M2 |
| | | M0 vs. M2-7b | | 3 | 1.817434 | 0.61115584 | M0 |
| BMP7 | Site model | M0 vs. M3 | | 4 | 569.175784 | 0 | M3 |
| | | M1a vs. M2a | | 2 | 0 | 1.00000000 | M1a |
| | | M7 vs. M8 | | 2 | 1.294466 | 0.52350955 | M7 |
| | Branch-site model | MA vs. null-7a | 559 T 0.959* | 1 | 12.347524 | 0.00044158 | MA |
| | | MA vs. null-7b | | 1 | 0 | 1.00000000 | Null |
| | Branch model | M0 vs. M1 | | 18 | 67.531604 | 0.00000012 | M1 |
| | | M0 vs. M2 | | 1 | 4.203038 | 0.04035253 | M2 |
| BMP9 | Site model | M0 vs. M3 | | 4 | 697.961866 | 0 | M3 |
| | | M1a vs. M2a | | 2 | 0 | 1.00000000 | M2a |
| | Branch-site model | M7 vs. M8 | | 2 | 19.615106 | 0.00005504 | M8 |
| | | MA vs. null | 26 F 0.996** | 1 | 22.64134 | 0.00000195 | MA |

*($p < 0.05$); **($p < 0.01$)

## Temporal expression patterns of bone morphogenetic protein genes in *Omobranchus punctatus*

The temporal expression patterns of BMP genes during different developmental stages of *O. punctatus* were analyzed by real-time qRT–PCR. The mRNA abundances of different BMP genes in five developmental stages were compared. BMP8 was the most abundantly expressed BMP gene during *O. punctatus* development. Notably, the relative expression levels of all BMP genes were significantly upregulated from 5 to 30 dph and peaked at 30 dph, with the exception of BMP9. Moreover, the lowest relative expression of all genes was observed at 5 dph. As shown in Figure 10, most of the BMP genes, especially BMP2, BMP3b-1, BMP3b-2, BMP4, and BMP11-1, were abundantly expressed at 30 dph, when the beak-like teeth were conical teeth that had not yet healed and only the tips of the teeth were ossified. The expression of BMPs significantly increased during the period from 16 to 30 dph, which indicates that osteoblasts are rapidly generated at this stage and that activation of ossification in *O. punctatus* occurs in preparation for the sclerosing of bones and beak-like teeth. The expression of BMP genes across development can be divided into three patterns. BMP2, BMP3b-1, BMP4, BMP11-1, BMP11-2, BMP11-3, and BMP16 exhibited the first expression pattern (Figure 9A): the expression levels increased from 5 to 30 ph, with the highest expression at

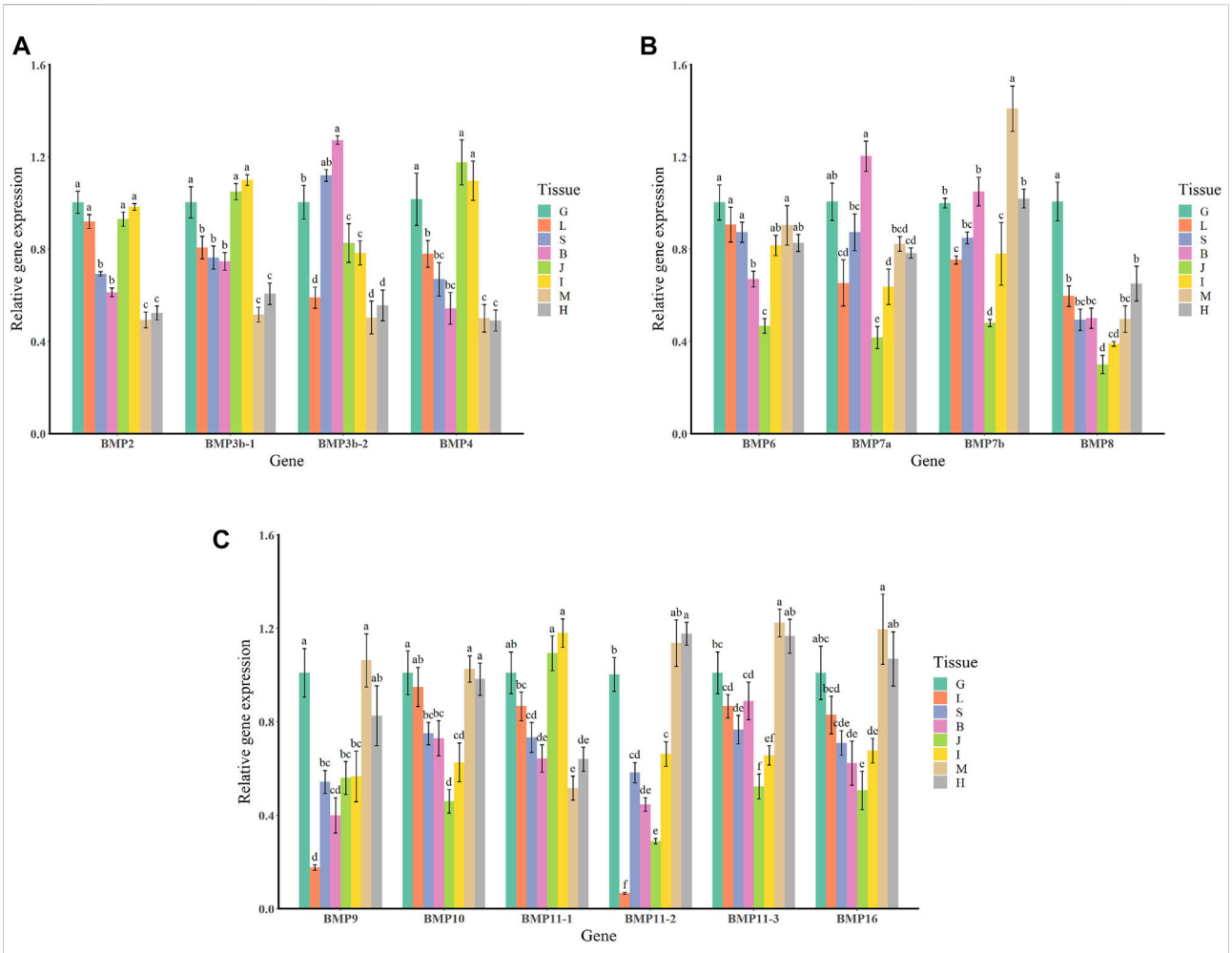



FIGURE 8
Relative expression levels of BMP genes in selected tissues of *O. punctatus*. **(A)** Shows the tissue-specific expression patterns of BMP2, BMP3b-1, BMP3b-2 and BMP4; **(B)** shows the tissue-specific expression patterns of BMP6, BMP7a, BMP7b and BMP8; **(C)** shows the tissue-specific expression patterns of BMP9, BMP10, BMP11-1, BMP11-2, BMP11-3 and BMP16. G represents the gills; L represents the liver; S represents the spleen; B represents the brain; J represents the surrounding tissue of the beak-like teeth; I represents the intestine; M represents the muscle; H represents the heart. Different lowercase letters above the standard error bars indicate significant differences in expression.

30 dph, and decreased sequentially from 30 to 60 dph, but the relative expression at 60 dph was still higher than the starting value (5 dph). In addition, the relative mRNA expression levels of BMP2, BMP3b-1, BMP11-1, BMP11-2, and BMP11-3 showed similar trends at the different developmental stages of *O. punctatus*. However, the relative mRNA expression levels of BMP4 and BMP16 were significantly higher than those of the above five genes. The second expression pattern (Figure 9B) was different from the first, with the expression levels of BMP3b-2, BMP6, BMP7a, BMP7b, BMP8, and BMP10 increasing from 50 to 60 dph. Compared with other BMP genes, BMP7a showed no significant change in mRNA abundance from 5 to 60 dph. In addition, BMP8 showed a significant increase in mRNA expression at 30–60 dph. For the third expression pattern (Figure 9C), there was a continuous and subtle trend of increasing BMP9 mRNA expression from 5 to 50 dph, with the expression decreasing to a minimum at 60 dph.

## Discussion

To deeply explore the physiological role of *Oplegnathus* BMPs in jaw formation and development, systematic homology comparison was performed to identify the sequences of BMPs and determine their expression patterns. Phylogenetic analysis showed that the BMP11 subfamily of *O. fasciatus* and *O. punctatus* contained three members, BMP11-1, BMP11-2, and BMP11-3. The clustering results for BMP11-1 were consistent with the classification results of other species,

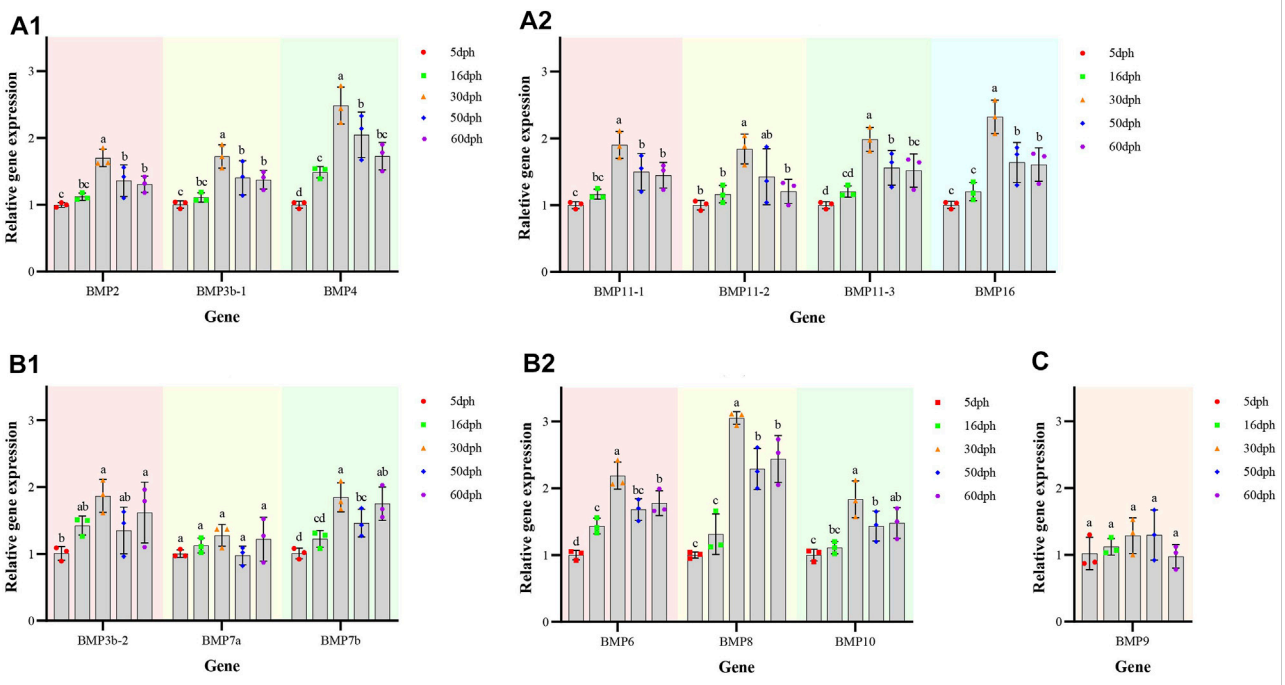



FIGURE 9
Relative expression levels of BMP genes in selected developmental stages of *O. punctatus*. **(A1,A2)** Represent temporal expression pattern one; **(A1)** shows the temporal expression pattern of BMP2, BMP3b-1 and BMP4; **(A2)** shows the temporal expression pattern of BMP11-1, BMP11-2, BMP11-3, and BMP16. **(B1,B2)** Represent temporal expression pattern two; **(B1)** shows the temporal expression pattern of BMP3b-2, BMP7a and BMP7b; **(B2)** shows the temporal expression pattern of BMP6, BMP8 and BMP10. **(C)** Represents temporal expression pattern three, the temporal expression pattern of BMP9.

while BMP11-2 and BMP11-3 clustered individually. According to the tissue-specific expression pattern of the BMP11 subfamily in *O. punctatus*, BMP11-2 was similar to BMP11-3, and its expression was upregulated in heart and muscle tissues. In contrast, BMP11-1 was downregulated in heart and muscle tissues and upregulated in intestinal tissue and the epithelial tissue and mesenchymal tissue of the beak-like teeth. Findings in mice suggest that BMP11 specifies positional identity along the anterior/posterior axis (McPherron et al., 1999). According to the clustering results, it is speculated that the BMP11-1 in these fishes functions similarly to that in mice and plays an important role in the development of the skeleton and teeth. This also explains the upregulated expression in the tissues around the beak-like teeth of *O. punctatus* shown in Figure 8. In contrast, the downregulated expression of BMP11-2 and BMP11-3 in the surrounding tissues of beak-like teeth suggested weak roles in the formation of the beak-like teeth of *O. punctatus*. The functions of BMP11-2 and BMP11-3 need to be further explored.

The expression changes of target genes during the healing and ossification of *O. punctatus* beak-like teeth were monitored by qRT–PCR. Notably, with the exception of BMP9, the expression of which was not obvious, all BMP genes tested were upregulated at 30 dph. It is inferred that osteoblasts are synthesized rapidly at approximately 30 days of age and enter the period of rapid ossification. For evidence of ossification, we can refer to the study of ossification in juvenile turbot. The parasphenoid bone of the cranial element is the first to undergo ossification beginning at 19 dph in juvenile turbot. Subsequently, maxilla, premaxilla, dentary, frontal, opercular, and preopercular ossification begin. Ossification of the intracranial skeletal system was visible when the juveniles reached full length at 25 dph. The degree of ossification was more obvious at 35 dph (Lv, 2018). As shown in Figure 10, according to the results of alizarin red bone staining, *O. punctatus* had a tapered distribution of teeth at 45 dph, and most of the teeth were stained blue with bromophenol blue, showing a cartilaginous state; only the tips of the teeth were red, indicating that ossification was initiated at the tips of the teeth. Similar findings were reported in pufferfish (Thiery et al., 2017). After 50 dph, discretely arranged tapered teeth showed a tendency to heal, and the teeth were red, indicating that ossification was already complete. This suggests that ossification is a biological outcome that requires the accumulation of osteoblasts before it can occur, and the BMP/Smads pathway is involved in osteoblast differentiation (Ding et al., 2021), while the RT–PCR results showed that most

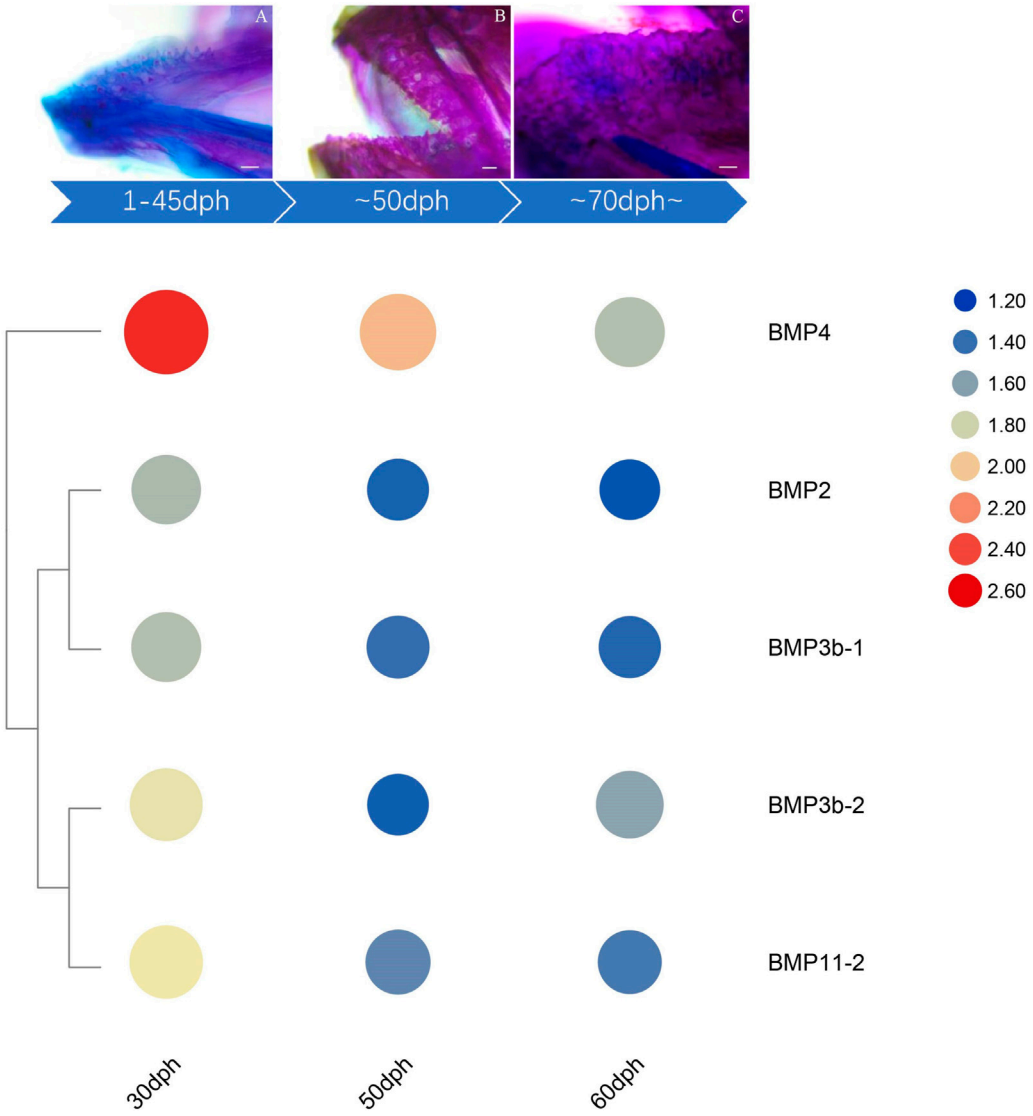



FIGURE 10
Expression of beak-like teeth-specific BMP genes at critical nodes in beak-like teeth development of *O. punctatus*. The upper panel indicates the results of alizarin red bone staining at critical time of beak-like teeth development, and the lower panel of heatmap indicates the chronological expression of BMP genes that are highly expressed in the surrounding tissue of the beak-like teeth.

of the BMP genes, especially BMP2, BMP3b-1, BMP3b-2, BMP4, and BMP11-1, were abundantly expressed at 30 dph, presumably preparing the beak-like teeth for healing and sclerosis at 45–50 dph. Compared with that of other genes, the expression of BMP4, BMP8, and BMP16 maintained high levels throughout the five developmental stages of *O. punctatus*, suggesting that they all participate in beak-like tooth growth. Studies have proven that BMP4 can be expressed in the upper beak mesenchyme of Darwin's finches and is strongly correlated with the regulation of bird beak shape. With increased BMP4 expression, the birds can grow deeper and longer beaks (Abzhanov et al., 2004; Campas et al., 2010). In addition, similar findings have been reported in fish, where the BMP4 gene was involved in the regulation of mandarin fish jaw remodeling (Cao et al., 2020). Studies have also shown that allelic variation and expression of BMP4 are closely related to cichlid jaw morphogenesis. BMP4 may be the main factor responsible for the diversity of jaws in different vertebrate classes (birds and fishes) (Albertson and Kocher, 2006). Based on the upregulated expression of BMP4 in five different developmental stages of *O. punctatus*, this gene may also play a regulatory role in the development of its beak-like teeth. This also explains the high expression of the

BMP4 gene in the epithelial tissue and mesenchymal tissue of the beak-like teeth shown in Figure 8. Thus, BMP4 represents a candidate gene for the adaptive evolution of beak-like tooth healing in *O. punctatus*. Studies in zebrafish show that BMP signaling is activated during caudal fin regeneration, and BMP8a expression is upregulated after caudal fin amputation (Schebesta et al., 2006). On the basis of our results that the expression of BMP8 is upregulated at 30 dph and the total expression level is higher than that of other genes, we infer that BMP8 plays an important role in the induction of bone and tooth formation.

Phylogenetic analysis revealed three BMP3 members in *O. fasciatus*: OfBMP3a, OfBMP3b-1, and OfBMP3b-2. *O. punctatus* contains two BMP3 members: OpBMP3b-1 and OpBMP3b-2. The OfBMP3a of *O. fasciatus* was selected as the foreground branch, and the BMP3s of the other species were selected as the background branch for positive selection pressure analysis. The results revealed 11 positively selected amino acid sites. In addition, the expression of BMP3b was upregulated in the epithelial tissue and mesenchymal tissue of the beak-like teeth, as shown in Figure 8. In a study of canine tooth root development, it was shown that BMP3 plays a role in the development of the tooth germ and tooth root and participates in the differentiation of dental follicle cells into cementoblasts and periodontal cells (Xuan et al., 2006). However, most studies suggest that BMP3 functions as an antagonist of BMPs, while BMP3 synthesis is generally dependent on BMPs, suggesting that a feedback mechanism is required to maintain the balance between BMPs and their antagonists. This does not conflict with the result that BMP3 is positively selected (Canalis et al., 2003). Therefore, similar to the role of BMP3 in the development of intermuscular bone in blunt snout bream (Zhang et al., 2018), we speculate that BMP3 may act as a negative regulator of skeletal growth, similarly to other positive BMPs, during the formation and development of the beak-like teeth and maintain internal stability. Evolutionary selection pressure analysis revealed a positively selected site in BMP7 under the branch-site model. Positive selection pressure on these loci may be an important reason for the tissue-specific expression and functional diversity of BMP genes. Studies have shown that BMP7 is an important initiation signal molecule for tooth germ development. It is secreted by epithelial cells and induces the expression of mesenchyme. In addition, BMP7 can participate in the formation of tooth germ and mediate the differentiation of mesenchymal cells (Tasli et al., 2014; Gao et al., 2015). Furthermore, studies on mouse teeth suggest that BMP2, BMP4, and BMP7 play a role in regulating tooth eruption and shape development and may be involved in the induction and formation of dentin and enamel (Aberg et al., 1997). Therefore, it is inferred that BMP7 plays an important role in the development of the special beak-like teeth of *O. fasciatus* and *O. punctatus*, which distinguish them from other species evolutionarily.

## Conclusion

In summary, we identified a total of 19 BMP genes in the *O. fasciatus* genome, and 16 members of the BMP gene family were identified in *O. punctatus,* with BMP1, BMP12 and BMP14 being lost. Subsequently, to gain a comprehensive and deep understanding of the BMP gene family and its distribution in the genome, we performed phylogenetic analysis, BMP sequence alignment, motif and domain composition analysis, collinearity analysis, and selection pressure analysis, among other analyses. We also described the expression patterns of BMPs in the gills, liver, spleen, brain, the surrounding tissues of beak-like teeth teeth, intestine, muscle, and heart of *O. punctatus*. Most of the BMP genes were widely expressed. The results on the temporal regulation of these BMP genes during tooth development in *O. punctatus* are notable. The expression of BMPs significantly increased from 16 to 30 dph, which indicates that osteoblasts are rapidly generated at 30 dph and that activation of osteoblasts in *O. punctatus* occurs in preparation for ossification of the bones and beak-like teeth at 40–50 dph. Our investigation revealed the effects of the BMP gene family on the healing and ossification development of the beak-like teeth of *O. fasciatus* and *O. punctatus* at the levels of genome evolution and RNA molecules, providing an immeasurable reference for future research on the molecular biology, physiology, and evolution of *Oplegnathus*. However, the detailed functions and regulatory mechanisms of each BMP gene still require further exploration.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

## Ethics statement

The animal study was reviewed and approved by the Experimental Animal Ethics Committee of the Institute of Oceanology, Chinese Academy of Sciences.

## Author contributions

YX and JL conceived and designed the project and revised the manuscript. YM and YX performed the genomic investigations and wrote the manuscript. ZX, YW, HZ, GG, LW, TW, and NZ participated in data analysis, discussion and figure preparation. All authors read and approve the final manuscript.

## Funding

This work was supported by grants from the National Key Research and Development Program (2018YFD0901204), Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (GML2019ZD0402), Key Deployment Projects of Center for Ocean Mega-Science, Chinese Academy of Sciences (Frontier Cross-category, COMS2020Q05),China Agriculture Research System (CARS-47), Major Agricultural Application Technology Innovation Project of Shandong Province (SD2019YY01), STS project (KFZD-SW-106, ZSSD-019, 2017T3017, and KFJ-STS-QYZX-020), Qingdao National Laboratory for Marine Science and Technology (2018SDKJ0502-2 and 2015ASKJ02), National Natural Science Foundation of China (No. 31672672).

## Conflict of interest

Author ZX was employed by the company Weihai Haohuigan Marine Biotechnology Co.

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

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.938473/full#supplementary-material

## References

Aberg, T., Wozney, J., and Thesleff, I. (1997). Expression patterns of bone morphogenetic proteins (Bmps) in the developing mouse tooth suggest roles in morphogenesis and cell differentiation. *Dev. Dyn.* 210 (4), 383–396. doi:10.1002/(SICI)1097-0177(199712)210:4<383:AID-AJA3>3.0.CO;2-C

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., and Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science* 305 (5689), 1462–1465. doi:10.1126/science.1098095

Albertson, R. C., and Kocher, T. D. (2006). Genetic and developmental basis of cichlid trophic diversity. *Heredity* 97 (3), 211–221. doi:10.1038/sj.hdy.6800864

Andreucci, R. D., Britski, H. A., and Carneiro, J. (1982). Structure and evolution of tetraodontoid teeth - an autoradiographic study (pisces, tetraodontiformes). *J. Morphol.* 171 (3), 283–292. doi:10.1002/jmor.1051710304

Asharani, P. V., Keupp, K., Semler, O., Wang, W. S., Li, Y., Thiele, H., et al. (2012). Attenuated BMP1 function compromises osteogenesis, leading to bone fragility in humans and zebrafish. *Am. J. Hum. Genet.* 90 (4), 661–674. doi:10.1016/j.ajhg.2012.02.026

Bai, Y. L., Gong, J., Zhou, Z. X., Li, B. J., Zhao, J., Ke, Q. Z., et al. (2021). Chromosome-level assembly of the southern rock bream (*Oplegnathus fasciatus*) genome using PacBio and hi-C technologies. *Front. Genet.* 12, 811798. doi:10.3389/fgene.2021.811798

Bielawski, J. P., and Yang, Z. (2003). Maximum likelihood methods for detecting adaptive evolution after gene duplication. *J. Struct. Funct. Genomics* 3 (1-4), 201–212. doi:10.1023/a:1022642807731

Bond, J. S., and Beynon, R. J. (1995). The astacin family of metalloendopeptidases. *Protein Sci.* 4 (7), 1247–1261. doi:10.1002/pro.5560040701

Bragdon, B., Moseychuk, O., Saldanha, S., King, D., Julian, J., Nohe, A., et al. (2011). Bone morphogenetic proteins: A critical review. *Cell. Signal.* 23 (4), 609–620. doi:10.1016/j.cellsig.2010.10.003

Campas, O., Mallarino, R., Herrel, A., Abzhanov, A., and Brenner, M. P. (2010). Scaling and shear transformations capture beak shape variation in Darwin's finches. *Proc. Natl. Acad. Sci. U. S. A.* 107 (8), 3356–3360. doi:10.1073/pnas.0911575107

Canalis, E., Economides, A. N., and Gazzerro, E. (2003). Bone morphogenetic proteins, their antagonists, and the skeleton. *Endocr. Rev.* 24 (2), 218–235. doi:10.1210/er.2002-0023

Canalis, E. (2009). Growth factor control of bone mass. *J. Cell. Biochem.* 108 (4), 769–777. doi:10.1002/jcb.22322

Cao, X., Ma, C., Xia, X., Zhang, R., Chen, X., and Zhao, J. (2020). Cloning and expression analysis of jaw remodeling developmental gene Bmp4 in *Siniperca chuatsi*. *Genomics Appl. Biol.* 39 (11), 5075–5083. doi:10.13417/j.gab.039.005075

Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi:10.1016/j.molp.2020.06.009

Chen, H. Y., Shi, S., Acosta, L., Li, W. M., Lu, J., Bao, S. D., et al. (2004). BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. *Development* 131 (9), 2219–2231. doi:10.1242/dev.01094

Chen, L., Dong, C., Kong, S., Zhang, J., Li, X., Xu, P., et al. (2017). Genome wide identification, phylogeny, and expression of bone morphogenetic protein genes in tetraploidized common carp (*Cyprinus carpio*). *Gene* 627, 157–163. doi:10.1016/j.gene.2017.06.020

Cheng, H. W., Jiang, W., Phillips, F. M., Haydon, R. C., Peng, Y., Zhou, L., et al. (2003). Osteogenic activity of the fourteen types of human bone morphogenetic proteins (BMPs). *J. Bone Jt. Surg. Am.* 85A (8), 1544–1552. doi:10.2106/00004623-200308000-00017

Daluiski, A., Engstrand, T., Bahamonde, M. E., Gamer, L. W., Agius, E., Stevenson, S. L., et al. (2001). Bone morphogenetic protein-3 is a negative regulator of bone density. *Nat. Genet.* 27 (1), 84–88. doi:10.1038/83810

Dendooven, A., van Oostrom, O., van der Giezen, D. M., Leeuwis, J. W., Snijckers, C., Joles, J. A., et al. (2011). Loss of endogenous bone morphogenetic protein-6 aggravates renal fibrosis. *Am. J. Pathol.* 178 (3), 1069–1079. doi:10.1016/j.ajpath.2010.12.005

Desroches-Castan, A., Tillet, E., Ricard, N., Ouarne, M., Mallet, C., Belmudes, L., et al. (2019). Bone morphogenetic protein 9 is a paracrine factor controlling liver sinusoidal endothelial cell fenestration and protecting against hepatic fibrosis. *Hepatology* 70 (4), 1392–1408. doi:10.1002/hep.30655

Ding, W., Cui, Y., and Zhang, L. (2021). Signal pathways regulating bone formation. *Chin. J. Biochem. Mol. Biol.* 37 (04), 428–436. doi:10.13865/j.cnki.cjbmb.2021.01.1392

Dingerkus, G., and Uhler, L. D. (1977). Enzyme clearing of alcian blue stained whole small vertebrates for demonstration of cartilage. *Stain Technol.* 52 (4), 229–232. doi:10.3109/10520297709116780

Ducy, P., and Karsenty, G. (2000). The family of bone morphogenetic proteins. *Kidney Int.* 57 (6), 2207–2214. doi:10.1046/j.1523-1755.2000.00081.x

Edgar, R. C. (2004). Muscle: A multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinforma.* 5, 113. doi:10.1186/1471-2105-5-113

Galloway, S. M., McNatty, K. P., Cambridge, L. M., Laitinen, M. P. E., Juengel, J. L., Jokiranta, T. S., et al. (2000). Mutations in an oocyte-derived growth factor

gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nat. Genet.* 25 (3), 279–283. doi:10.1038/77033

Gamer, L. W., Wolfman, N. M., Celeste, A. J., Hattersley, G., Hewick, R., Rosen, V., et al. (1999). A novel BMP expressed in developing mouse limb, spinal cord, and tail bud is a potent mesoderm inducer in Xenopus embryos. *Dev. Biol.* 208 (1), 222–232. doi:10.1006/dbio.1998.9191

Gao, B., Zhou, X., Zhou, X. D., Pi, C. X., Xu, R. S., Wan, M., et al. (2015). BMP7 and EREG contribute to the inductive potential of dental mesenchyme. *Sci. Rep.* 5, 9903. doi:10.1038/srep09903

Geng, L. Y., Zhang, C. S., and Jiang, Y. L. (2010). Molecular cloning and expression analysis of porcine bone morphogenetic protein 11 (BMP11) gene. *J. Anim. Vet. Adv.* 9 (23), 2986–2989. doi:10.3923/javaa.2010.2986.2989

Goncalves, A., and Zeller, R. (2011). Genetic analysis reveals an unexpected role of BMP7 in initiation of ureteric bud outgrowth in mouse embryos. *Plos One* 6 (4), e19370. doi:10.1371/journal.pone.0019370

Gong, J., Li, B. J., Zhao, J., Zhou, Z. X., Ke, Q. Z., Zhu, Q. H., et al. (2022a). Sex-specific genomic region identification and molecular sex marker development of rock bream (*Oplegnathus fasciatus*). *Mar. Biotechnol.* 24 (1), 163–173. doi:10.1007/s10126-022-10095-2

Gong, J., Zhao, J., Ke, Q. Z., Li, B. J., Zhou, Z. X., Wang, J. Y., et al. (2022b). First genomic prediction and genome-wide association for complex growth-related traits in Rock Bream (*Oplegnathus fasciatus*). *Evol. Appl.* 14, 523–536. doi:10.1111/eva.13218

Hotten, G. C., Matsumoto, T., Kimura, M., Bechtold, R. F., Kron, R., Ohara, T., et al. (1996). Recombinant human growth/differentiation factor 5 stimulates mesenchyme aggregation and chondrogenesis responsible for the skeletal development of limbs. *Growth factors.* 13 (1-2), 65–74. doi:10.3109/08977199609034567

Kakizawa, Y., Kamishikiryo, K., Shirato, M., Maehara, S., Fujii, H., Iesato, S., et al. (1980). The tooth development of the parrot perch, *Oplegnathus fasciatus*, (family Oplegnathidae, Teleostei). *J. Nihon Univ. Sch. Dent.* 22 (4), 211–226. doi:10.2334/josnusd1959.22.211

Kim, R. Y., Robertson, E. J., and Solloway, M. J. (2001). Bmp6 and Bmp7 are required for cushion formation and septation in the developing mouse heart. *Dev. Biol.* 235 (2), 449–466. doi:10.1006/dbio.2001.0284

King, J. A., Marker, P. C., Seung, K. J., and Kingsley, D. M. (1994). BMP5 and the molecular, skeletal, and soft-tissue alterations in short ear mice. *Dev. Biol.* 166 (1), 112–122. doi:10.1006/dbio.1994.1300

Kosa, J. P., Kis, A., Bacsi, K., Balla, B., Nagy, Z., Takacs, I., et al. (2011). The protective role of bone morphogenetic protein-8 in the glucocorticoid-induced apoptosis on bone cells. *Bone* 48 (5), 1052–1057. doi:10.1016/j.bone.2011.01.017

Li, M., Zhang, R., Fan, G. Y., Xu, W. T., Zhou, Q., Wang, L., et al. (2021). Reconstruction of the origin of a neo-Y sex chromosome and its evolution in the spotted knifejaw, Oplegnathus punctatus. *Mol. Biol. Evol.* 38 (6), 2615–2626. doi:10.1093/molbev/msab056

Li, Q., Gu, X., Weng, H. L., Ghafoory, S., Liu, Y., Feng, T., et al. (2013). Bone morphogenetic protein-9 induces epithelial to mesenchymal transition in hepatocellular carcinoma cells. *Cancer Sci.* 104 (3), 398–408. doi:10.1111/cas.12093

Liu, Z., and Lian, W. (2018). Molecular mechanisms of bone morphogenetic protein, Wnt, fibroblast growth factor and sonic hedgehog signaling pathways in tooth development. *Chin. J. Tissue Eng. Res.* 22 (36), 5897–5904. doi:10.3969/j.issn.2095-4344.1000

Lv, X. (2018). *Growth characteristic, osteological ontogeny and deformity in larval and juvenile* Scophthalmus maximus *and Epinephelus lanceolatus*. Master's Thesis. Qingdao (Shangdong): University of Chinese Academy of Sciences.

Marques, C. L., Fernandez, I., Viegas, M. N., Cox, C. J., Martel, P., Rosa, J., et al. (2016). Comparative analysis of zebrafish bone morphogenetic proteins 2, 4 and 16: Molecular and evolutionary perspectives. *Cell. Mol. Life Sci.* 73 (4), 841–857. doi:10.1007/s00018-015-2024-x

Maschette, D., Fromont, J., Platell, M. E., Coulson, P. G., Tweedley, J. R., Potter, I. C., et al. (2020). Characteristics and implications of spongivory in the Knifejaw Oplegnathus woodwardi (Waite) in temperate mesophotic waters. *J. Sea Res.* 157, 101847. doi:10.1016/j.seares.2020.101847

Mazerbourg, S., and Hsueh, A. J. W. (2006). Genomic analyses facilitate identification of receptors and signalling pathways for growth differentiation factor 9 and related orphan bone morphogenetic protein/growth differentiation factor ligands. *Hum. Reprod. Update* 12 (4), 373–383. doi:10.1093/humupd/dml014

McPherron, A. C., Lawler, A. M., and Lee, S. J. (1999). Regulation of anterior posterior patterning of the axial skeleton by growth differentiation factor 11. *Nat. Genet.* 22 (3), 260–264. doi:10.1038/10320

Patino, L. C., Walton, K. L., Mueller, T. D., Johnson, K. E., Stocker, W., Richani, D., et al. (2017). BMP15 mutations associated with primary ovarian insufficiency reduce expression, activity, or synergy with GDF9. *J. Clin. Endocrinol. Metab.* 102 (3), 1009–1019. doi:10.1210/jc.2016-3503

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29 (9), e45. doi:10.1093/nar/29.9.e45

Rafael, M. S., Laize, V., and Cancela, M. L. (2006). Identification of *Sparus aurata* bone morphogenetic protein 2: Molecular cloning, gene expression and *in silico* analysis of protein conserved features in vertebrates. *Bone* 39 (6), 1373–1381. doi:10.1016/j.bone.2006.06.021

Schebesta, M., Lien, C. L., Engel, F. B., and Keating, M. T. (2006). Transcriptional profiling of caudal fin regeneration in zebrafish. *ScientificWorldJournal.* 6, 38–54. doi:10.1100/tsw.2006.326

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11 molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38 (7), 3022–3027. doi:10.1093/molbev/msab120

Tasli, P. N., Aydin, S., Yalvac, M. E., and Sahin, F. (2014). Bmp 2 and bmp 7 induce odonto- and osteogenesis of human tooth germ stem cells. *Appl. Biochem. Biotechnol.* 172 (6), 3016–3025. doi:10.1007/s12010-013-0706-0

Thiery, A. P., Shono, T., Kurokawa, D., Britz, R., Johanson, Z., Fraser, G. J., et al. (2017). Spatially restricted dental regeneration drives pufferfish beak development. *Proc. Natl. Acad. Sci. U. S. A.* 114 (22), E4425–E4434. doi:10.1073/pnas.1702909114

Tsumaki, N., Tanaka, K., Arikawa-Hirasawa, E., Nakase, T., Kimura, T., Thomas, J. T., et al. (1999). Role of CDMP-1 in skeletal morphogenesis: Promotion of mesenchymal cell recruitment and chondrocyte differentiation. *J. Cell. Biol.* 144 (1), 161–173. doi:10.1083/jcb.144.1.161

Turingan, R. G. (1994). Ecomorphological relationships among caribbean tetraodontiform fishes. *J. Zoology* 233, 493–521. doi:10.1111/j.1469-7998.1994.tb05279.x

von Bubnoff, A., and Cho, K. W. Y. (2001). Intracellular BMP signaling regulation in vertebrates: Pathway or network? *Dev. Biol.* 239 (1), 1–14. doi:10.1006/dbio.2001.0388

Wang, S., Xu, W. T., Li, M., Wang, J., Wang, L., Zhai, J. M., et al. (2020). Molecular cloning and expression pattern analysis of TGF-β1 in spotted knifejaw (Oplegnathus punctatus). *Prog. Fish. Sci.* 41 (3), 78–87. doi:10.19663/j.issn2095-9869.20190309001

Wolfman, N. M., Hattersley, G., Cox, K., Celeste, A. J., Nelson, R., Yamaji, N., et al. (1997). Ectopic induction of tendon and ligament in rats by growth and differentiation factors 5, 6, and 7, members of the TGF-beta gene family. *J. Clin. Invest.* 100 (2), 321–330. doi:10.1172/jci119537

Wozney, J. M., Rosen, V., Celeste, A. J., Mitsock, L. M., Whitters, M. J., Kriz, R. W., et al. (1988). Novel regulators of bone-formation - molecular clones and activities. *Science* 242 (4885), 1528–1534. doi:10.1126/science.3201241

Xiao, Y. S., Xiao, Z. Z., Ma, D. Y., Liu, J., and Li, J. (2019). Genome sequence of the barred knifejaw *Oplegnathus fasciatus* (temminck & schlegel, 1844): The first chromosome-level draft genome in the family Oplegnathidae. *Gigascience* 8 (3), giz013. doi:10.1093/gigascience/giz013

Xiao, Y., Xiao, Z., Ma, D., Zhao, C., Liu, L., Wu, H., et al. (2020). Chromosome-level genome reveals the origin of neo-Y chromosome in the male barred knifejaw *Oplegnathus fasciatus*. *iScience* 23 (4), 101039. doi:10.1016/j.isci.2020.101039

Xie, X.-D., Zhao, L., Wu, Y.-F., and Wang, J. (2020). Role of bone morphogenetic protein 1/tolloid proteinase family in the development of teeth and bone. *Hua xi kou qiang yi xue za zhi = Huaxi kouqiang yixue zazhi = West China J. stomatology* 38 (5), 589–593. doi:10.7518/hxkq.2020.05.020

Xu, P., Zhang, X. F., Wang, X. M., Li, J. T., Liu, G. M., Kuang, Y. Y., et al. (2014). Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat. Genet.* 46 (11), 1212–1219. doi:10.1038/ng.3098

Xuan, K., Yang, F., and Wen, L. (2006). Expression of bmp3, bmp4 and bmp7 during dog permanent tooth root formation. *J. Pract. Stomatology* 22 (6), 807–810. doi:10.3969/j.issn.1001-3733.2006.06.018

Yang, Z. H. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15 (5), 568–573. doi:10.1093/oxfordjournals.molbev.a025957

Yang, Z. (2007). Paml 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24 (8), 1586–1591. doi:10.1093/molbev/msm088

Zhang, J. Z., Nielsen, R., and Yang, Z. H. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22 (12), 2472–2479. doi:10.1093/molbev/msi237

Zhang, W. Z., Lan, T., Nie, C. H., Guan, N. N., and Gao, Z. X. (2018). Characterization and spatiotemporal expression analysis of nine bone morphogenetic protein family genes during intermuscular bone development in blunt snout bream. *Gene* 642, 116–124. doi:10.1016/j.gene.2017.11.027

Zhou, X., Tao, Y., Liang, C., Zhang, Y., Li, H., Chen, Q., et al. (2015). BMP3 alone and together with TGF-beta promote the differentiation of human mesenchymal stem cells into a nucleus pulposus-like phenotype. *Int. J. Mol. Sci.* 16 (9), 20344–20359. doi:10.3390/ijms160920344

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 18 July 2022
doi: 10.3389/fgene.2022.941340

# Development of a High-Density 665 K SNP Array for Rainbow Trout Genome-Wide Genotyping

*Maria Bernard[1,2], Audrey Dehaullon[1], Guangtu Gao[3], Katy Paul[1], Henri Lagarde[1], Mathieu Charles[1,2], Martin Prchal[4], Jeanne Danon[5], Lydia Jaffrelo[5], Charles Poncet[5], Pierre Patrice[6], Pierrick Haffray[6], Edwige Quillet[1], Mathilde Dupont-Nivet[1], Yniv Palti[3], Delphine Lallias[1] and Florence Phocas[1]**

[1]INRAE, AgroParisTech, GABI, Université Paris-Saclay, Jouy-en-Josas, France, [2]INRAE, SIGENAE, Jouy-en-Josas, France, [3]USDA, REE, ARS, NEA, NCCCWA, Kearneysville, WV, United States, [4]South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Faculty of Fisheries and Protection of Waters, University of South Bohemia, Vodňany, Czechia, [5]INRAE-UCA, Plateforme Gentyane, UMR GDEC, Clermont-Ferrand, France, [6]SYSAAF, Campus de Beaulieu, Rennes, France

OPEN ACCESS

**Edited by:**
*Nguyen Hong Nguyen, University of the Sunshine Coast, Australia*

**Reviewed by:**
*James W. Kijas, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia*
*Tereza Manousaki, Hellenic Centre for Marine Research (HCMR), Greece*

***Correspondence:**
*Florence Phocas*
*florence.phocas@inrae.fr*

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

**Received:** *11 May 2022*
**Accepted:** *24 June 2022*
**Published:** *18 July 2022*

**Citation:**
*Bernard M, Dehaullon A, Gao G, Paul K, Lagarde H, Charles M, Prchal M, Danon J, Jaffrelo L, Poncet C, Patrice P, Haffray P, Quillet E, Dupont-Nivet M, Palti Y, Lallias D and Phocas F (2022) Development of a High-Density 665 K SNP Array for Rainbow Trout Genome-Wide Genotyping. Front. Genet. 13:941340. doi: 10.3389/fgene.2022.941340*

Single nucleotide polymorphism (SNP) arrays, also named « SNP chips », enable very large numbers of individuals to be genotyped at a targeted set of thousands of genome-wide identified markers. We used preexisting variant datasets from USDA, a French commercial line and 30X-coverage whole genome sequencing of INRAE isogenic lines to develop an Affymetrix 665 K SNP array (HD chip) for rainbow trout. In total, we identified 32,372,492 SNPs that were polymorphic in the USDA or INRAE databases. A subset of identified SNPs were selected for inclusion on the chip, prioritizing SNPs whose flanking sequence uniquely aligned to the Swanson reference genome, with homogenous repartition over the genome and the highest Minimum Allele Frequency in both USDA and French databases. Of the 664,531 SNPs which passed the Affymetrix quality filters and were manufactured on the HD chip, 65.3% and 60.9% passed filtering metrics and were polymorphic in two other distinct French commercial populations in which, respectively, 288 and 175 sampled fish were genotyped. Only 576,118 SNPs mapped uniquely on both Swanson and Arlee reference genomes, and 12,071 SNPs did not map at all on the Arlee reference genome. Among those 576,118 SNPs, 38,948 SNPs were kept from the commercially available medium-density 57 K SNP chip. We demonstrate the utility of the HD chip by describing the high rates of linkage disequilibrium at 2–10 kb in the rainbow trout genome in comparison to the linkage disequilibrium observed at 50–100 kb which are usual distances between markers of the medium-density chip.

**Keywords: SNP, single nucleotide polymorphism, sequence, high-density chip, linkage disequilibrium, rainbow trout, doubled haploid lines, isogenic lines**

## INTRODUCTION

Next-generation sequencing (NGS) has transformed the fields of quantitative, ecological and evolutionary genetics by enabling the discovery and cost-effective genotyping of thousands to millions of variants across the genome, allowing for genome-wide association studies (GWAS) of complex traits, genomic selection (GS) through accurate inference of relationships among individuals (Meuwissen and Goddard, 2010), inbreeding (Kardos et al., 2015), population

structure, and genetic diversity studies. Large numbers of densely genotyped individuals are required to get accurate results thanks to a high SNP density along the genome that constructs strong linkage disequilibrium between SNP and causative mutations (de Roos et al., 2008). However, regardless of the animal or plant species, it remains very challenging to cost-effectively genotype large numbers of individuals at polymorphic sites in all the genomes. An appealing strategy is to use a cheaper and reduced-density SNP chip with markers being chosen for optimizing the imputation accuracy to higher density genotypes. Genotype imputation describes the process of predicting genotypes that are not directly assayed in a sample of individuals (Marchini and Howie, 2010). Imputation has become a standard practice in research to increase genome coverage and improve GS accuracy and GWAS resolution, as a large number of samples can be genotyped at lower density (and lower cost) then imputed up to denser marker panels or to sequence level, using information from a limited reference population (Phocas, 2022).

Two main methods are employed for large-scale and genome-wide SNP genotyping. Array-based methods use flanking probe sequences to interrogate pre-identified SNPs (often named "SNP chips"). The alternative genotyping-by-sequencing (GBS) methods call SNPs directly from the genome (Davey et al., 2011). In GBS methods, either restriction enzymes are used to target sequencing resources on a limited number of cut sites (Baird et al., 2008) or low-coverage whole genome resequencing is performed. Low-coverage GBS followed by imputation has been proposed as a cost-effective genotyping approach for human genetics studies (Pasaniuc et al., 2012), as well as farmed species (Gorjanc et al., 2017) that cannot afford a high development of genomic tools. Nevertheless, compared to GBS methods, SNP chips offer a robust and easily replicable way of genotyping samples at a consistent set of SNPs, with very low rates of missing data.

Medium (~thousands to tens of thousands of loci) and high (~hundreds of thousands of loci) density SNP chips have been routinely developed for commercial species to perform genomic selection (Meuwissen et al., 2001) and to identify genes playing significant roles in livestock and crop performances (Goddard et al., 2016). SNP chips developed for model organisms or farmed species have also been utilised to address evolutionary and conservation questions, in particular in animal populations. For example, they have been used to identify signatures of adaptation in cattle (Gautier et al., 2010) or genes under selection in grey wolves (Schweizer et al., 2016), characterize the genetic diversity and inbreeding levels in pig (Silió et al., 2013), sheep (Mastrangelo et al., 2014), cattle (Rodríguez-Ramilo et al., 2015) or fish (D'Ambrosio et al., 2019), and infer the genomic basis of recombination rate variation in cattle (Sandor et al., 2012) or sheep (Johnston et al., 2016; Petit et al., 2017).

While there is now over ten fish and shellfish species for which commercial SNP arrays had been developed (Boudry et al., 2021), most of those contain only about 50 to 60 K SNPs. Such medium-density chips are sufficient for genomic selection purposes but are clearly too low-density tools for fine QTL detection and help in identification of causal variants. As rainbow trout (*Oncorhynchus mykiss*) is a major academic model for a wide range of investigations in disciplines such as cancer research, toxicology, immunology, physiology, nutrition, developmental or evolutionary biology in addition to quantitative genetics and breeding (Thorgaard et al., 2002), it is important to get access to very high-density genomic tools for this salmonid species.

For rainbow trout, SNP discovery has been firstly done through sequencing of restriction-site associated DNA (RAD) libraries (Palti et al., 2014), reduced representation libraries (RRL) (Sánchez et al., 2009), and RNA sequencing (Sánchez et al., 2011). A first commercial medium-density Axiom® Trout Genotyping array (hereafter termed 57 K chip) has then been developed (Palti et al., 2015) and produced by Affymetrix (Thermofisher). Since then it has been largely used in population genetics studies (Larson et al., 2018; D'Ambrosio et al., 2019; Paul et al., 2021), GWAS and GS accuracy works for various traits in farmed populations (Gonzalez-Pena et al., 2016; Vallejo et al., 2017; Vallejo et al., 2019; Reis Neto et al., 2019; Rodríguez et al., 2019; Yoshida et al., 2019; Fraslin et al., 2019; Fraslin et al., 2020; Karami et al., 2020; D'Ambrosio et al., 2020; Blay et al., 2021a; Blay et al., 2021b). However, out of the 57,501 SNPs included in this chip, nearly 20,000 were found to be unusable because they were either duplicated due to the ancestral genome duplication or showing primer polymorphism in five French commercial or experimental lines (D'Ambrosio et al., 2019). Of the 57,501 markers from the original chip, 50,820 are uniquely localized on the Swanson reference genome (Pearse et al., 2019), and in the remaining number, only 38,332 markers pass the control quality filters [no primer polymorphism, call rate > 97%, Minor Allele Frequency (MAF) >0.001 over 3,000 fish from five French lines].

To overcome these limitations as well as to get access to a more powerful tool for GWAS and population genetics studies in rainbow trout, the aim of our study was to develop a high-density SNP array. To develop this resource for rainbow trout, we made use of a large set of resequencing data from 31 doubled haploid (DH) lines from Washington State University (WSU) and Institut National de la Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE). In the United States, 12 WSU DH lines have been created by androgenesis (Young et al., 1996) while in France 19 DH INRAE lines (called isogenic lines) were produced by gynogenesis (Quillet et al., 2007). The 12 WSU DH lines as well as seven of the INRAE isogenic lines served as basic material for the variant search and SNP selection for the 57 K chip (Palti et al., 2015).

In this study, we describe how we overcame the limitations of duplications in the rainbow trout genome, in order to identify and locate polymorphisms. We describe the subset of detected SNPs that was selected for inclusion on a custom high-density SNP chip. It was used to genotype 463 samples from two different French commercial populations. We test the genotyping success rates, that is, the proportion of SNPs included on the array that are polymorphic and successfully genotyped. We demonstrate the utility of this SNP chip to infer linkage disequilibrium in the genome of this species.

# MATERIALS AND METHODS

## Use of the USDA Database for Initial SNP Detection

Gao et al. (2018) constituted a first large SNP database (USDA1) by performing high coverage whole genome resequencing (WGS) with 61 unrelated samples, representing a wide range of rainbow trout and steelhead populations. Of the 61 samples, 11 were doubled-haploid lines from Washington State University (WSU), 12 were aquaculture samples from AquaGen (Norway), 38 were from wild and hatchery populations from a wide range of geographic distribution (California, Oregon, Washington and Idaho states in the United States; Canada; Kamtchatka Penninsula in Russia). Overall, 31,441,105 SNPs were identified with 30,302,087 SNPs located on one of the 29 chromosomes of the Swanson reference genome assembly (Omyk_1.0; GenBank, assembly accession GCA_002163495.1) (Pearse et al., 2019).

A second database (USDA2) with 17,889,078 SNPs coming from resequencing of 24 USDA samples was added to the initial USDA1 database. The samples were composed of 12 representatives from the USDA-NCCCWA odd-year class and 12 from the even-year class as previously described (NCBI BioProject PRJNA681179; Liu et al., 2021). The SNP discovery analysis followed the methods of (Gao et al., 2018).

By merging these two databases using BCFtools (Danecek et al., 2021), we constituted a single USDA database that contained 35,732,342 distinct SNPs, with 34,170,401 placed on the 29 chromosomes or mitochondrial chromosome of the Swanson reference genome. SNP filtering was performed to remove non bi-allelic variants and SNPs with MAF <1% using a vcf_filter.py homemade script (**Supplementary File S1**). The final USDA clean database contained 29,024,315 SNPs.

## Whole Genome Resequencing of INRAE Isogenic Lines and Use of the INRAE Database for SNP Detection

Genomic DNA was extracted from fin clips of 19 rainbow trout INRAE isogenic lines. Whole-genome paired-end sequencing libraries were prepared and sequenced using the Illumina HiSeq 2000, Hi Seq 3000 or HiSeq X-Ten platforms at a depth of genome coverage ranging from 10X to 32X per sample. The 19 isogenic lines were sequenced in two batches that were processed successively. The first batch contained sequencing data from 12 samples (doubled haploid individuals) coming from 11 isogenic lines. The second batch contained sequencing data from 17 samples (doubled haploid individuals) coming from 17 isogenic lines (9 lines already sequenced in batch 1; and 8 lines not previously sequenced). Overall, 10 out of the 19 isogenic lines were sequenced twice. This resulted in a total of 8,911,630,867 paired reads with a median of 321,575,464 per sample.

Sequence reads from each of the 12 samples from the first batch were mapped to the Swanson rainbow trout reference genome (GenBank assembly accession GCA_002163495.1; Pearse et al., 2019) using BWA MEM v.0.7.12 (Li, 2013). We then ran Samtools sort [v1.3.1, (Danecek et al., 2021)] to sort the alignment data by chromosome and scaffold locations. Afterwards, PCR duplicates were marked using Picard Tools (v.2.1.1, Broad Institute, 2019) MarkDuplicates. Variant calling was then performed for each sample using GATK (v3.7; McKenna et al., 2010) HaplotypeCaller (options *-stand_call_conf 30 -mbq 10*), leading to 12 vcf files. A variant reference file containing 1,207,861 high quality SNPs was generated by keeping variants with $QUAL \geq 1{,}050$ from the vcf files. This file was then used for the recalibration step, using GATK BaseRecalibrator and PrintReads. The recalibrated BAM files were then used as input for the variant calling step using GATK HaplotypeCaller in ERC GVCF mode. The resulting 12 GVCF files were then merged into a single vcf file containing 24,944,575 variants using GATK GenotypeGVCFs. The vcf file was then filtered as follows using GATK VariantFiltration: $DP < 120$; $MQ < 30.0$; $QUAL < 600$; $AN<12$. To filter out putative PSVs (Paralogous Sequence Variants), we filtered out variants with heterozygous genotypes in at least two of the 12 doubled haploid samples. The filtered vcf file from the first batch contained 11,113,836 variants.

The second samples sequence batch were analyzed following the same procedure as for the first batch with few updates. Prior to sequence alignment, sequences have been filtered using trimmomatic 0.36 (Bolger et al., 2014) to remove Illumina Truseq adapters, trim low quality bases, keep trimmed reads with a sufficient length and average quality. These parameters removed 3.8% of the reads, keeping 6,349,173,142 reads over the 17 samples. Alignment software was updated to use BWA MEM v.0.7.15. First calling to create a high quality variants set to recalibrate the BAM files was avoided by directly using the final vcf file from the first batch analysis. These recalibrated BAM have been submitted to GATK Haplotype caller as before to generate GVCF files. To increase confidence in the SNP calling, we also added 2 other SNP callers: Samtools mpileup and FreeBayes 1.1.0 (Garrison and Marth, 2012). GATK calling results were jointly genotyped using GATK GenotypeGVCFs on the 12 GVCF files from the first batch and the 17 newly generated GVCF files. This calling procedure resulted in 3 VCF files, one for each caller. Calling from GATK contained 29 samples (from the 19 isogenic lines, i.e., with 10 lines replicated) and 31,454,943 variants; Freebayes and Mpileup was used only on the second batch and contained 19 samples and 25,805,271 and 30,340,281 variants respectively.

The final step for variant calling was to intersect the 3 calling datasets using VCFtools_0.1.12a (Danecek et al., 2011), to keep only variants called by the 3 callers (genotypes kept were the GATK ones). SNP and INDEL were separated using GATK SelectVariants, and SNP were filtered with GATK VariantFiltration by following the GATK recommendations ($QD < 2.0 \parallel MQ < 40.0 \parallel FS > 60.0 \parallel SOR > 3.0 \parallel MQRankSum < -12.5 \parallel ReadPosRankSum < -8.0$). This constitutes the INRAE1 variants dataset which includes 14,439,713 SNPs.

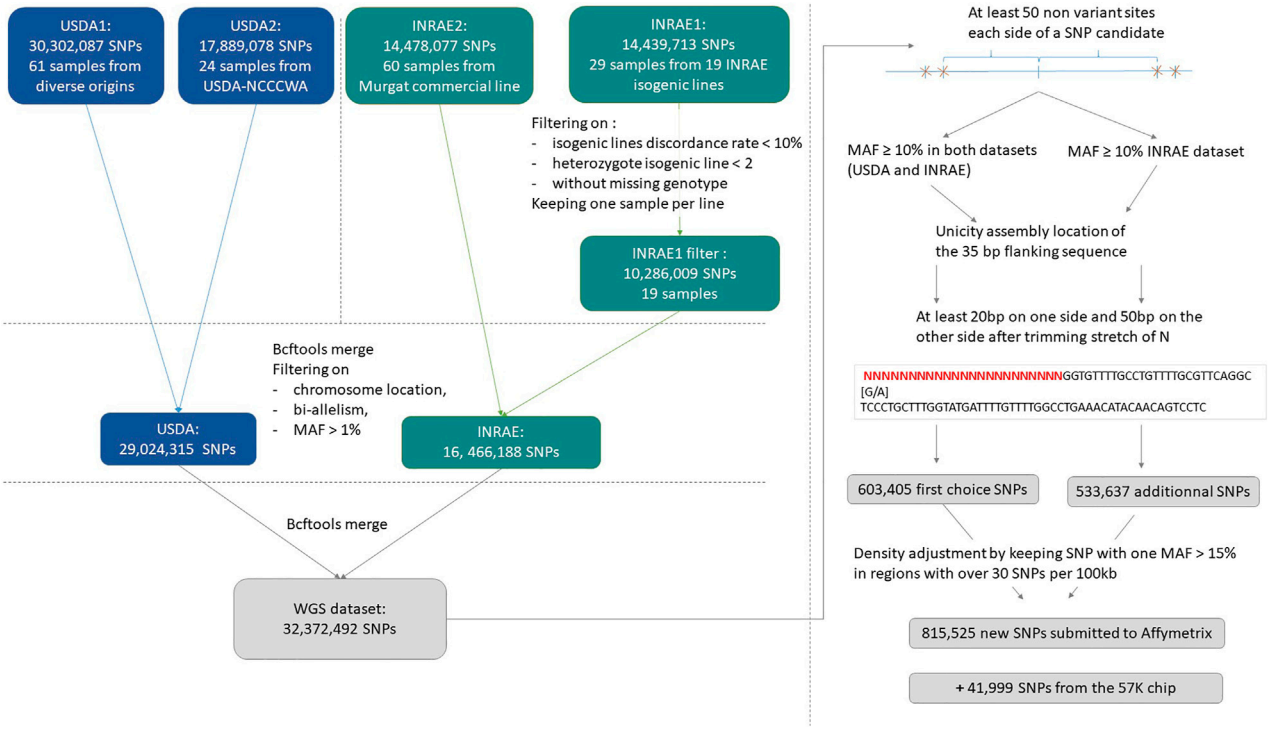



**FIGURE 1 |** Process for submitted SNPs for inclusion on a high-density genotyping array.

Using a vcf_filter.py homemade script to parse VCF file, INRAE1 dataset was filtered to keep only bi-allelic SNP localized on the 29 trout chromosomes or mitochondrial chromosome, fully genotyped for all 29 samples. As 10 isogenic lines were duplicated, we also checked genotype consistency and removed SNP with more than 1 isogenic line genotype discordance. Finally, we kept one sample per isogenic line (with the deepest sequencing) and filtered out SNP with more than 1 heterozygote genotype as they may represent duplicated genome regions. Among the 14,439,713 variants, we kept 10,286,009 SNPs (71.23%).

We merged them using BCFtools with a second dataset INRAE2, containing 14,478,077 SNPs called from 60 samples of a commercial line from "Les Fils de Charles Murgat" (Beaurepaire, France) and whose resequencing was described in Fraslin et al. (2020).

This merged dataset was filtered like the merged USDA dataset, to keep bi-allelic SNP localized on the 29 trout chromosomes of the Swanson reference genome or mitochondrial chromosome, with a MAF >1%. The final INRAE cleaned database contained 16,466,188 SNPs.

## Merging the USDA and INRAE SNP Databases and SNP Preselection

A total of 32,372,492 distinct SNPs were selected for consideration for the HD chip, from a combination (BCFtools merge) of the USDA and INRAE databases (https://doi.org/10.5281/zenodo.6657091).

An overview of the process to detect and select SNPs for inclusion on the array is provided in **Figure 1**.

SNPs were further filtered to be at least 50 base pairs from the closest identified SNP which resulted in a subset of 3,679,547 SNPs.

During vcf files merging, additional alleles may be added on shared variant positions and some variants previously removed from the INRAE dataset (on replicate discordance or high isogenic heterozygosity rate) may be reincluded. Thus, for a first SNP preselection, in addition to filtering SNPs with MAF ≥10% in both the USDA and INRAE databases, we applied filters on bi-allelism variant and on a maximum number of 4 heterozygote INRAE isogenic lines.

Assessment included a check for duplicate flanking information suggesting repetitive elements, and an assessment of the complexity of the flanking sequence:

1) Unicity of at least one side 35 bp-sequence for each SNP. This was done by blasting (default blastn parameters, Camacho et al., 2009) the 35 bp on the reference assembly genome and by checking that the best match was unique and located on the expected chromosome.
2) Trimming of each side 50 bp sequence if it contained more than 3 successive N. Variants were kept if at least the shortest trimmed sequence contained 20 bp and the other 50 bp (trimN.py homemade script, **Supplementary File S2**).

This first high-quality selection represented 633,405 SNPs. Trimmed flanking sequence each side of the SNP was extracted

for all SNPs and formatted for Affymetrix (Thermo Fisher Scientific, United States) according to their specifications.

From this first submission to Affymetrix quality control, only 457,086 SNPs were qualified as recommended to be designable for the HD array and among them, only 351,755 were not ambiguous, meaning they were not of the type [G/C] or [T/A] that would require 4 probes instead of only two to distinguish the alleles.

To get sufficient recommended variants and to avoid the selection of markers that will use twice the space used by the others on the HD array, we decided to resubmit a large second set of variants to Affymetrix quality check. The same procedure was applied to produce a second more relaxed preselected set of SNPs by keeping SNPs with a MAF ≥10% in the INRAE dataset only. This second preselection contained 533,637 additional SNPs. Among that additional set, 134,086 SNPs were specific to the INRAE dataset while the others were also present in the USDA dataset but with MAF below 10%.

We merged the first recommended set of 457,086 SNPs with this additional set of 533,637 SNPs. Then we removed all ambiguous SNPs of type [G/C], [C/G], [T/A] or [A/T]. Finally, densities were adjusted such that in regions with more than 30 SNPs retained per 100 kb by the previous filters, we only kept SNPs with MAF ≥15% in at least one of the two INRAE or USDA databases.

This procedure resulted in a selection of 815,525 SNPs for the final submission in October 2020 to Affymetrix for assessment of the suitability of the SNPs for inclusion on a custom AXIOM 96HT SNP chip. Of the submitted SNPs, a total of 623,544 SNPs were deemed to be "designable" (recommended or neutral) in either the forward or reverse flanking sequence based on the Affymetrix pconvert score.

## Keeping Informative Variants From the Medium-Density Axiom® Trout Genotyping Array

The INRAE and USDA research teams were willing to keep in the HD chip design the informative markers from the 57 K chip. Therefore 41,999 SNPs out of its 57,501 SNPs were designable in either forward and reverse directions and were kept for the HD chip design.

At the only exception of 8 specific SNPs, all the markers had a unique position on the Swanson reference genome and MAF > 5% in at least one French or North American population. Among them, 38,826 SNPs were also put on a 200 K chip that was built on 120 resequenced mostly "wild" genomes from over 40 locations from Russia, Alaska Canada down through Washington, Oregon and California (Ben Koop's personal communication).

## Selection of SNPs for the HD-Trout SNP Chip

In total 664,531 SNPs corresponding to 701,602 probesets (some SNPs were tiled in both directions as both their forward and reverse flanking sequence was assessed to be neutral) passed the Affymetrix final quality control to be designed on the custom HD Axiom array. Only 40,987 of the 41,999 SNPs from the 57 K chip remained on the HD final design.

Among the selected SNPs, 664,503 were mapped on the 29 chromosomes of the Swanson reference genome (**Figure 2**), while 28 were positioned on the mitochondrial genome.

Based on the Swanson reference genome mapping, the average (median) SNP density on the chromosomes was 293 (324) SNPs per Mb (**Figure 2**), with SNP density varying from 2 to 774 SNPs per Mb. The average (median) inter-marker distance was 2.9 kb (1.3 kb). Maximum inter-marker distance was 243 kb and only 5.4% of inter-marker distance was over 10 kb (0.02% over 100 kb).

## Single Nucleotide Polymorphism Mapping on the Arlee Reference Genome

Recently USDA/ARS (Gao et al., 2021) released a second reference genome assembly (USDA_OmykA_1.1; GenBank assembly accession GCA_013265735.3) for *Oncorhynchus mykiss* as long reads-based *de-novo* assembly for a second WSU DH line, named Arlee line, had been performed. Because Arlee lineage was closer from the INRAE isogenic lines (Palti et al., 2014) than Swanson lineage, it was decided to keep for further analysis only the SNPs that were mapped uniquely on one of the 32 chromosomes of this new reference genome.

Therefore, the 701,602 SNP probe flanking sequences of the HD chip were realigned to the new Arlee reference genome using BLAST (blastn with default task, i.e., megablast, and filtering options -evalue 1e-5 -perc_identity 95 -best_hit_overhang 0.1 -best_hit_score_edge 0.1).

## SNP Validation

Given that the objective of our work was to get a powerful genomic tool with a large number of SNPs useful in diverse rainbow trout populations, we included fish from two French commercial lines that were genetically distinct from the lines used from the SNP discovery step.

We used fin samples collected from "Bretagne Truite" (Plouigneau, France) and "Viviers de Sarrance" (Sarrance, France) commercial lines, hereafter named LB and LC lines respectively, that were sampled for the FEAMP project Hypotemp (n° P FEA470019FA1000016). Pieces of caudal fin sampled from 463 fish (288 from LB line and 175 from LC line) were sent to Gentyane genotyping platform (INRAE, Clermont-Ferrand, France) for DNA extraction using the DNAdvance kit from Beckman Coulter following manufacturer instructions and genotyping using the newly constructed HD SNP array.

The first round of quality control was done by ThermoFisher software AxiomAnalysisSuite™ with threshold values of 97% for SNP call rate and 90% for sample call rate. All the 288 individuals of the LB line passed the preliminary control, while 174 out of the 175 individuals from LC line passed the control quality. Following array hybridization and imaging, genotypes were called using default settings in the Axiom Analysis Suite software and exported from the software in PLINK (Purcell et al., 2007) format.

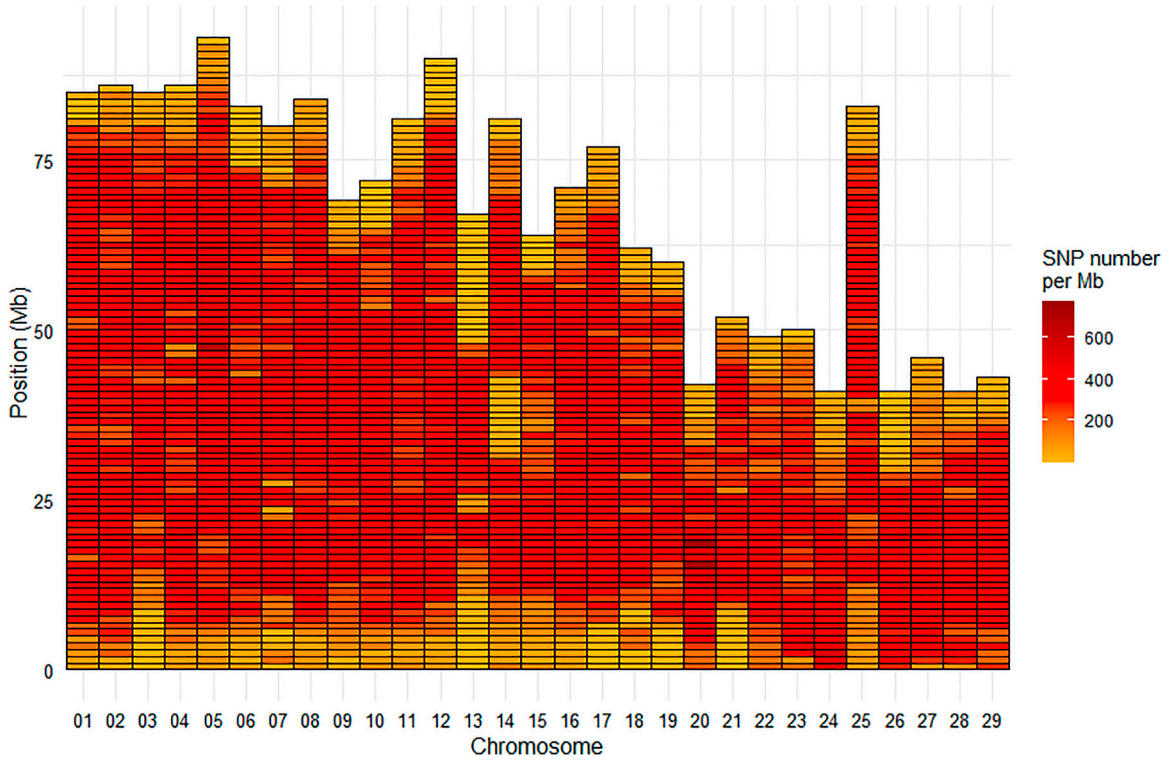



**FIGURE 2 |** Marker density per Mb for the HD Trout Affymetrix array with 664,503 SNPs positioned on the 29 chromosomes of the Swanson genome reference.

In addition, we used the WGS information of 20 samples sequenced in Gao et al. (2018)' study (with average genome coverage above ×20) to extract their genotypes for SNPs included in the HD chip and positioned on the Arlee reference genome (Gao et al., 2021; GCA_013265735.3). Those samples came from hatchery (Dworhak, L. Quinault, Quinault, Shamania) and wild (Elwha) populations from the North-West of United States (Four samples for each of the five populations) and were proved to be genetically close to each other and very distant from the Norwegian Aquagen aquaculture population (Gao et al., 2018). The idea was to infer and compare the level of linkage disequilibrium across the HD markers from wild/hatchery American populations and farmed French selected lines.

## Allele Frequencies and Linkage Disequilibrium Across Populations

We then used PLINK v1.9 (www.cog-genomics.org/plink2) to calculate allele frequencies, filter SNPs at low MAF or individuals with high identity by descent (IBD) values and derive linkage disequilibrium (LD) measured as the correlation coefficient r2, using the mapping of the SNP probe flanking sequences to the Arlee genome.

Allele frequencies were calculated per population for each SNP. SNPs were then filtered to only those with a MAF ≥5%, leaving 249,055 variants for American populations, 420,778 SNPs for the LB line, and 423,061 SNPs for the LC line. The set of individuals was also filtered using "*rel-cutoff 0.12*" to exclude one member of each pair of samples with observed genomic relatedness above 0.12, keeping 120 samples across populations, corresponding to 20, 45, and 55 individuals for American populations, LB and LC lines, respectively. Linkage disequilibrium (r2) between all pairs of SNPs on the same chromosome and at physical distances up to 1 Mb was then calculated using the PLINK options "*--r2 --ld-window 50000 –ld-window-kb 1001 –ld-window-r2 0.0*". The r2 values were binned into 2 kb units and per-bin averages calculated using R (R Core Team, 2019) for all chromosomes. The LD decay over physical distance up to 100 kb was then plotted in R.

# RESULTS

## SNP Identification and Characterization Based on the Swanson Reference Genome

Density of SNPs varied strongly from one chromosome to another with average SNP density per Mb ranging from 13,200 for Omy26 to 20,132 for Omy22. Across all chromosomes, the average SNP density per Mb was 16,483 SNPs (**Figure 3**). The Mb with the minimum density contained 451 SNPs while the Mb with the highest density contained 31,819 SNPs.

SNP identified in USDA or INRAE databases differed in terms of MAF distribution (**Figure 4**): 70% and 49% of SNPs had a MAF below 15% (40% and 15% had a MAF below 5%, respectively)

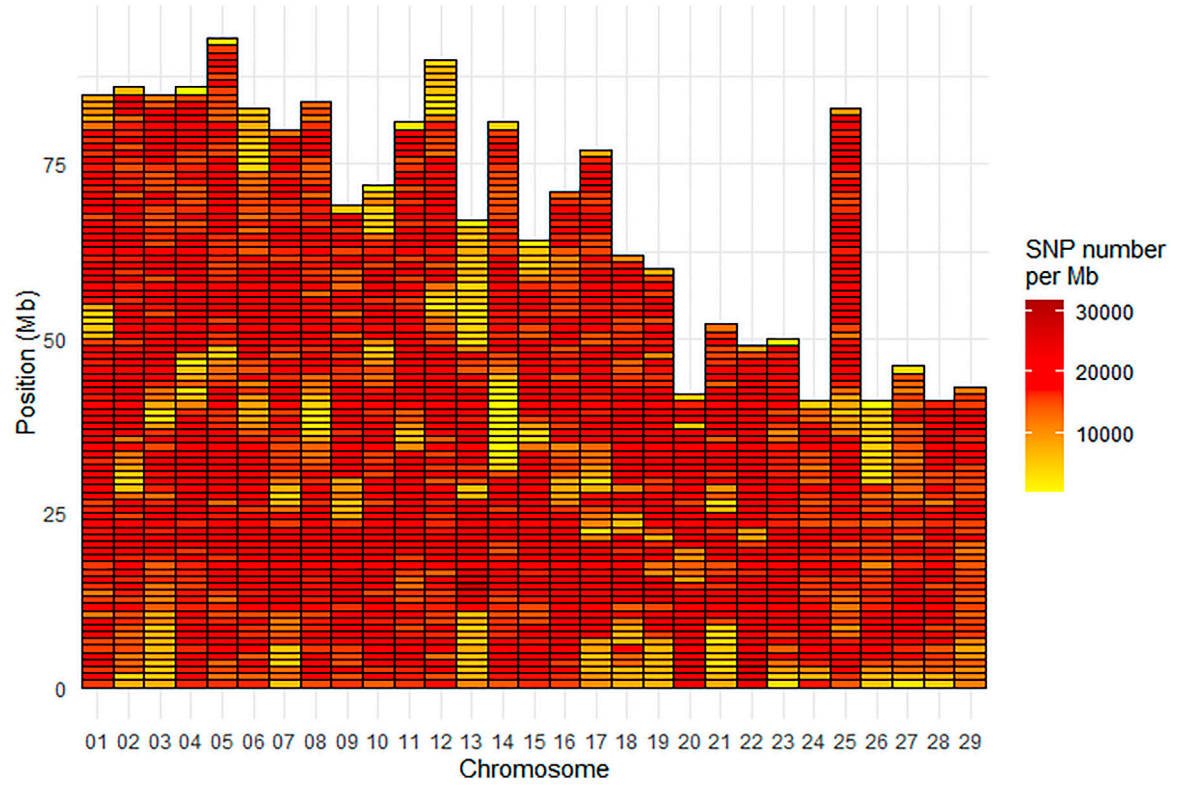



**FIGURE 3 |** SNP density per Mb for the INRAE_USDA full variant dataset (32.4M SNPs) located on the 29 chromosomes of the Swanson genome reference.

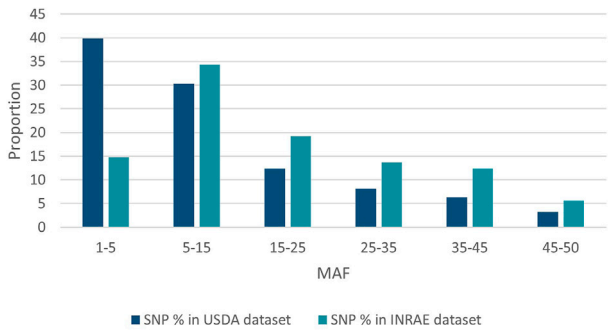



**FIGURE 4 |** MAF distribution of USDA or INRAE SNP datasets. These datasets have been filtered to keep bi-allelic SNP with a minimal MAF >1% in their respective populations.

while only 9.5% and 18% of SNPs had a MAF above 35% in the USDA and INRAE datasets respectively.

## HD Chip

Based on genotyping the 288 LB samples, 65.34% of markers were polymorphic, had individuals with all three genotypes, and passed Affymetrix filtering metrics in the Axiom Analysis Suite software to be categorized as "PolyHigh Resolution" variants. Of those that "failed" to be in that category, 15.35% passed filtering metrics but were monomorphic, 10.71% passed filtering metrics but the minor allele homozygote was missing, and the remainder 8.60% failed due to low call rates or other quality filters. The total number of best recommended markers was 91.81% corresponding to 610,115 SNPs out of the 664,531 genotyped variants.

Based on genotyping the 175 LC samples, 69.91% of markers were polymorphic, had individuals with all three genotypes, and passed Affymetrix filtering metrics in the Axiom Analysis Suite software to be categorized as "PolyHigh Resolution" variants. Of those that "failed" to be in that category, 5.63% passed filtering metrics but were monomorphic, 14.84% passed filtering metrics but the minor allele homozygote was missing, and the remainder 9.62% failed due to low call rates or other quality filters. The total number of best recommended markers was 90.86% corresponding to 603,768 SNPs out of the 664,531 genotyped variants.

Of the 664,531 SNPs which passed the Affymetrix quality filters and were included on the HD chip, 576,118 SNPs mapped uniquely on both reference genomes, and 12,071 SNPs did not map at all on the Arlee reference genome. **Supplementary Data S1** indicates both positions on the Swanson and Arlee reference genomes. Among those 576,118 SNPs, 38,948 SNPs were kept from the initial 57 K chip.

On the Arlee mapping (GCA_013265735.3), the average SNP density on the chromosomes was one SNP per 3.8 kb, or 266 SNPs per Mb. The median inter-marker distance was 1.5 kb with only 7% of the distances between successive

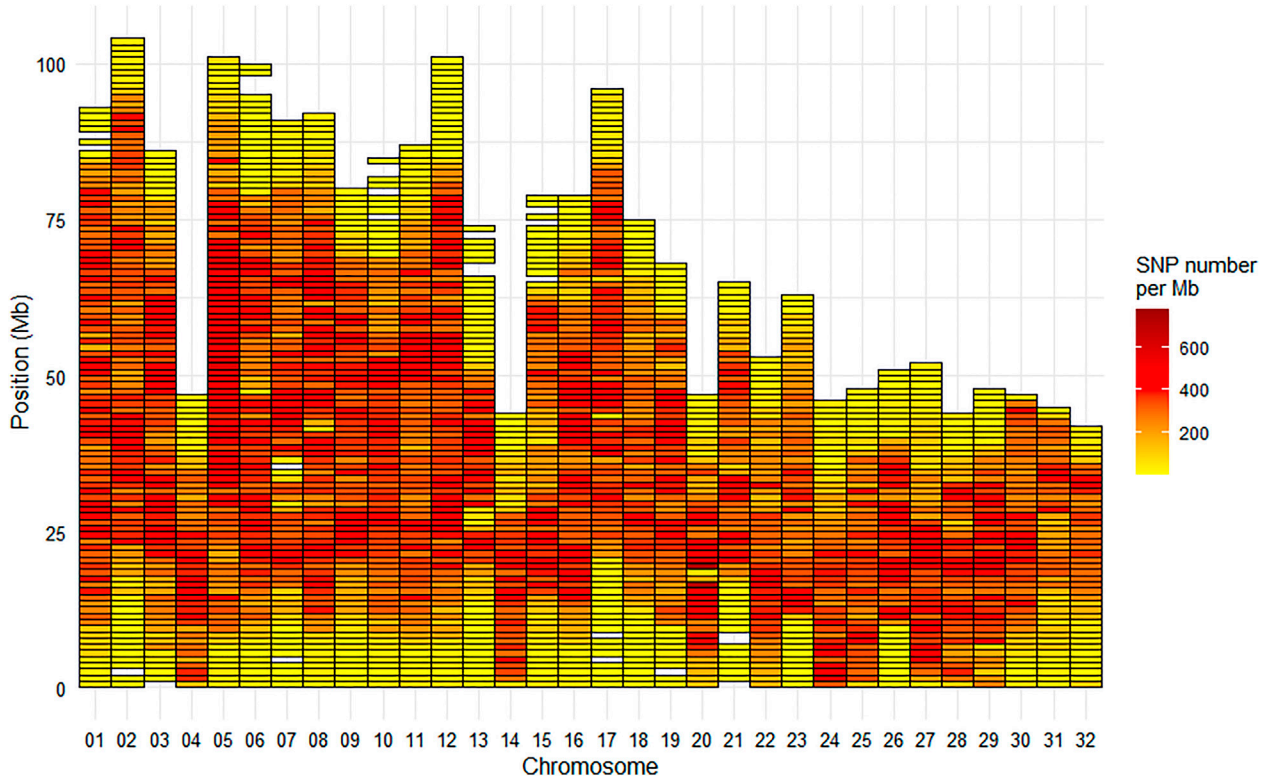



**FIGURE 5 |** Marker density per Mb for the HD Trout Affymetrix array with 576,118 SNPs positioned on the 32 chromosomes of the Arlee reference genome.

markers being above 10 kb. The largest gap was 4.16 Mb at the end of chromosome Omy6, the second largest gap was 2.94 Mb at the end of chromosome Omy10 and the third largest gap was 2.75 Mb at the end of chromosome Omy13 (**Figure 5**). Only five other gaps were above 2 Mb with values ranging from 2.3 to 2.5 Mb on chromosomes Omy7, Omy10, Omy15, and Omy21.

Finally, PLINK v1.9 software (www.cog-genomics.org/plink2) was used for a final SNP filtering based on keeping for further analysis SNPs with call rate above 95% and a deviation test from Hardy-Weinberg equilibrium (HWE) with a *p*-value < 10e-7 within each population. For LB line, 571,319 markers (474,937 being polymorphic) were kept after removing 1,136 miss genotyped SNPs and 3,663 ones with severe deviation from HWE. For LC line, 569,030 markers (487,940 being polymorphic) remained after removing 2,574 miss genotyped SNPs and 4,592 ones with severe deviation from HWE.

Regarding the American sequenced population, we extracted from the vcf files the genotypes for the 576,118 SNPs that were retained on the HD chip. Only 338,660 of those markers were polymorphic in the American population.

## MAF Distribution in the Two French HD Genotyped Populations

Compared to variants called from sequence data, the MAF distribution of the HD selected SNPs was skewed to common alleles (**Figure 6**) with over 70% of SNPs with MAF above 5% in each of the two populations, and over 20% of SNPs with MAF over 35% in both populations. Among polymorphic SNPs (MAF > 0.001), the average (median) MAF was 24.1% (23.6%) in the LB line and 23.0% (21.5%) in the LC line.

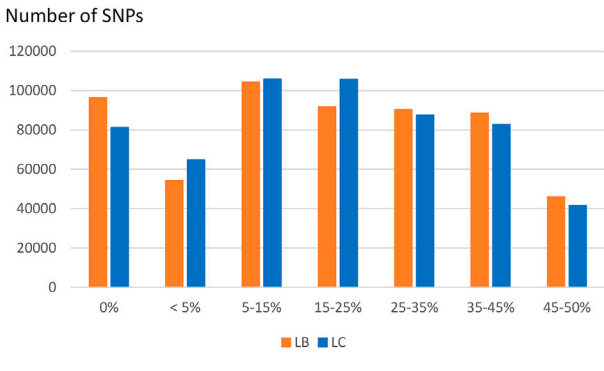



**FIGURE 6 |** Distribution of SNPs according to their MAF class in the LB and LC French commercial lines.

## Linkage Disequilibrium Analysis

The median intermarker distance was 2 kb and the corresponding average r2 between neighbouring markers was 0.47, 0.44, and 0.36 in LB, LC, and American population, respectively. As expected, average r2 tended to decrease with increasing

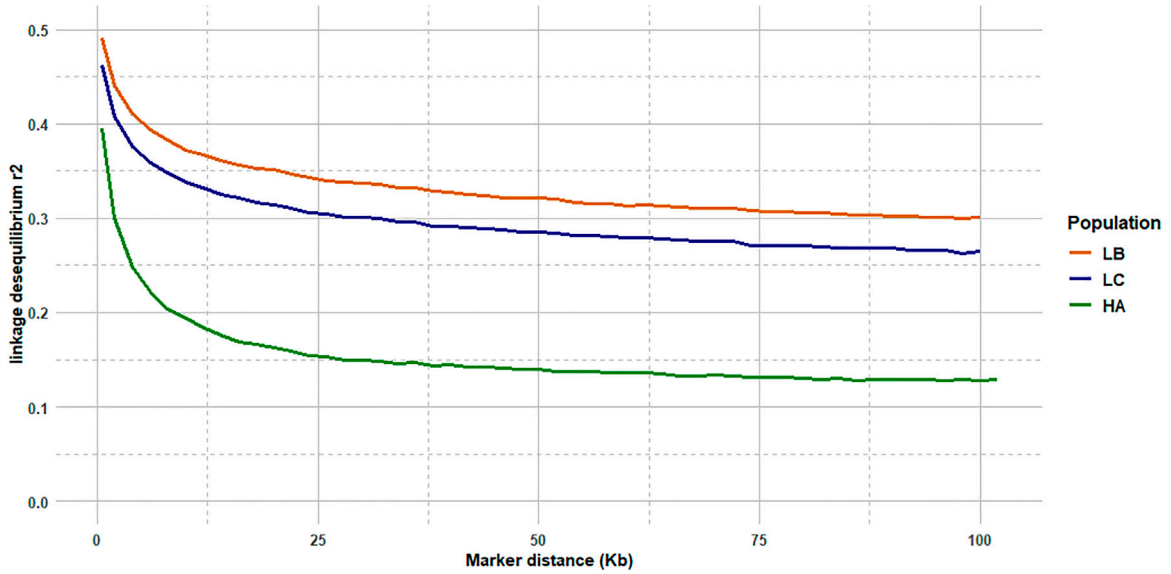



**FIGURE 7 |** LD decay from 2 to 100 kb intermarker distances (average over the 32 chromosomes) for the LB and LC French commercial lines and the HA American population.

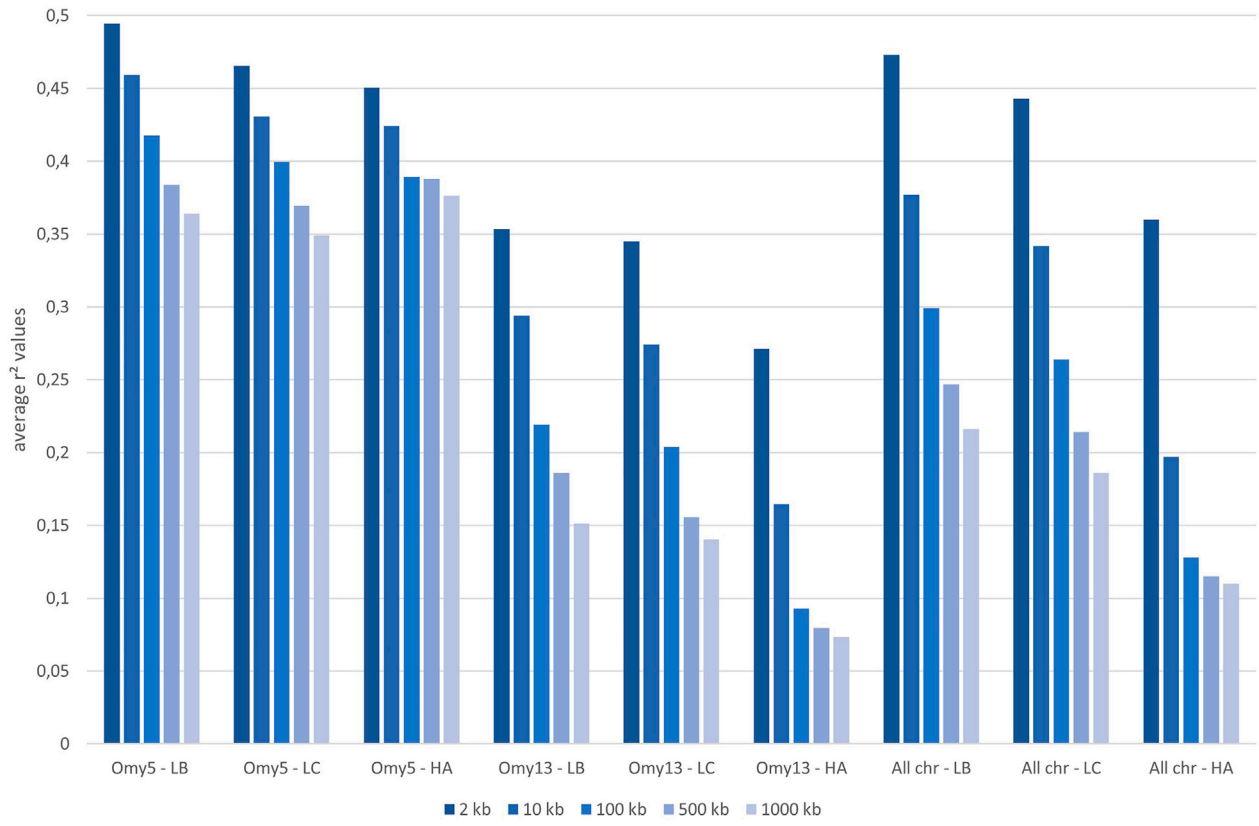



**FIGURE 8 |** Average linkage disequilibrium (r2 values) from 2 to 1,000 kb derived for all chromosomes and only for Omy5 or Omy13 in populations LB, LC, and HA, respectively.

distance between pairs of markers in all populations studied, the most rapid decline being over the first 10 kb (**Figure 7**). Linkage disequilibrium was very high, with r2 reaching 0.42, 0.39, and 0.27 at the average intermarker distance (4 kb) for LB, LC, and American population respectively; at 50 kb distance, r2 average values were 0.32, 0.29, and 0.14 (**Figure 7**). At 500 kb, values were 0.25, 0.21, and 0.18 and values were still 0.22, 0.19, and 0.11 at 1 Mb, respectively for LB, LC, and the American population (**Figure 8**). However, those r2 values may vary strongly from one chromosome to another as shown on **Figure 8** for chromosomes Omy5 and Omy13 with respectively higher and lower linkage disequilibrium observed in comparison to the average values derived for all chromosomes.

## DISCUSSION

In this study, based on the resequencing of tens of individuals from a diverse range of populations, we developed a high density (665K) SNP array that will be used for numerous applications, including genomic populations studies, GWAS or genomic selection. In fish, the first very high-density chip, named 930K XHD Ssal array, was developed for Atlantic Salmon using 29 fish from Aquagen lines and was a powerful tool to identify the key role of *VGLL3* gene on age at maturity (Barson et al., 2015) or the epithelial cadherin gene as the major determinant of the resistance of Atlantic salmon to IPNV (Moen et al., 2015). A similar approach was used in Atlantic salmon with whole genome re-sequencing of 20 fish from three diverse origins to generate a catalogue of 9.7M SNPs that were then filtered to design a 200K SNP chip (Yáñez et al., 2016). A similar number of 9.6M SNPs were identified for the development of a 700 K SNP chip in catfish (Zeng et al., 2017). Recently, a set of 82 fish were collected from six different locations of China and re-sequenced to identify 9.3M SNPs to design a 600K SNP chip for large yellow croaker (Zhou et al., 2020).

Based on the resequencing of 85 samples by USDA and 79 samples by INRAE, we identified 32,372,492 SNPs that were variants (MAF ≥ 1%) in either the USDA or the INRAE sets. More precisely, 29.0 and 16.4 million SNPs were identified in the USDA and INRAE datasets respectively for equivalent number of sequenced individuals. The higher number of SNPs detected in the USDA dataset probably resulted from the larger number of diverse populations included in the USDA dataset. The USDA database included 11 doubled haploid individuals and 50 individuals from 7 commercial, hatchery or wild populations, compared to the INRAE database that included 19 doubled haploid individuals derived from one experimental line and 60 individuals sampled from a single French commercial line. For comparison purposes, the influence of the numbers of sequenced individuals and populations or breeds on the number of identified SNPs can be exemplified in two large-scale projects, the 1000 human genomes project and the 1000 bull genomes project. In the human genome, a pilot phase identified ~15 million SNP based on the WGS of 179 individuals from four populations (Altshuler et al., 2010); increasing the number of sequences to 2,504 coming from 26 populations across the world increased considerably the number of identified SNPs to 84.7 million (Auton et al., 2015; Fairley et al., 2020). Similarly, the first phase of the 1000 bull genomes project identified 26.7 million SNPs based on the resequencing of 234 bulls from 3 breeds (Daetwyler et al., 2014); again, the number of SNPs increased to 84 million by sequencing 2,703 individuals from 121 breeds (Hayes and Daetwyler, 2019). Another study in chicken highlights the importance of sequencing a diverse set of individuals to identify a large catalogue of SNPs: WGS of 243 chickens from 24 chicken lines derived from diverse sources lead to the detection of about 139 million putative SNPs (Kranis et al., 2013).

In this study, the average distance between two successive variants was 60 bp, indicating important polymorphism level in the rainbow trout genome. This is consistent with the average SNP rate over all chromosomes of one SNP every 64 bp previously reported by Gao et al. (2018) in the Swanson rainbow trout reference genome. Such short average distance between successive variants was a strong limiting factor to preselect SNPs to design the HD chip. Indeed, an important technical issue in SNP array design is that very high SNP densities can potentially cause allele dropout when genotyping due to interferences between polymorphism at the marker position and at the probe designs that have to be monomorphic sequences flanking the marker candidates. When searching for markers with inter-marker distance over 50 bp that could be considered in the HD array design, we could only retain 3.68M SNPs.

Across all chromosomes, the average SNP density per Mb was 16,483 SNPs, i.e., slightly higher than the ~15,600 SNPs per Mb reported by Gao et al. (2018), although density of SNPs varied strongly from one chromosome to another (from 13,200 for Omy26 to 20,132 for Omy22). Interestingly, the lower SNP densities on Omy26 was also described in Gao et al. (2018) and associated with a higher proportion of SNPs being filtered out as putative paralogous sequence variants (PSV), as this chromosome shares high sequence homology with other chromosome arms in the genome as a result of delayed re-diploidization. Stronger variation in average SNPs density among chromosomes has been reported previously in chickens (Kranis et al., 2013) and humans (Zhao et al., 2003), with average value of 78 and 83.3 SNPs per kb across the genome but with some chromosomes having only 3 (on chromosome Z) and 2 (on chromosome Y) SNPs reported per kb respectively.

There was also a heterogeneous distribution of SNPs along the chromosomes, with a minimum density per Mb as low as 451 SNPs, and a maximum of 31,819 SNPs. Areas with less SNP density generally located at the telomeric parts or the centromeric parts (for metacentric chromosomes) of some chromosomes (e.g., Omy13 and Omy14) (**Figure 3**). Such heterogeneous distribution of SNPs has been previously reported in Eukaryotes, with potential explanations including heterogeneous recombination across the genome. It has been reported in a meta-analysis in eukaryotes that "heterogeneity in the distribution of crossover across the genome is a key determinant of heterogeneity in the distribution of genetic variation within and between populations" (Haenel et al., 2018). One broad-scale and general pattern observed within chromosomes is a lower recombination rate around centromeres

(Stapley et al., 2017) and higher rates at the telomeric parts (Sakamoto et al., 2000; Anderson et al., 2012). Because higher recombination rates are observed in telomeric than centromeric regions of chromosomes, a higher number of variants may be expected in the telomeres. However, in general, the telomeres have very long patterns of repeats which generate problems in reads mapping. In the centromeric regions, it is unclear whether or not suppressed recombination is linked to highly repetitive regions (Talbert and Henikoff, 2010). Last but not least, the complexity of the rainbow trout genome with its recent whole genome duplication and partial rediploidization, and patterns of tetrasomic inheritance (Pearse et al., 2019), can potentially explain the difficulties to sequence and assemble some parts of its genome and hence detect SNPs. In a recent paper, Gui et al. (2022) have reported several phenomena (such as massive sequence divergences, extensive chromosome rearrangements, large-scale transposon bursts) occurring during the polyploidization and rediploidization that could explain the difficulties in assembling the complex genomes of Salmonids and other tetraploid fish species. Indeed, rainbow trout has a high content (57.1%) of repetitive sequences (Pearse et al., 2019), similar to the 59.9% reported for Atlantic salmon (Lien et al., 2016).

Taking advantage of the biological characteristics of fish (external fertilization and embryonic development, viability of uniparental progeny), isogenic lines have been generated in some fish species (reviewed in Franěk et al., 2020), by either gynogenesis (Quillet et al., 2007) or androgenesis (Young et al., 1996) in rainbow trout. Both USDA and INRAE datasets included the sequencing of 11 and 19 doubled-haploid individuals respectively from 30 different isogenic lines. This number, quite large and unique in fish, makes it possible to take advantage of both the within-line characteristics (homozygosity, isogenicity) and between-line variability. In particular, rainbow trout isogenic lines are being used for the development of genomic tools: the trout genome is the result of a whole genome duplication event that occurred about 96 Mya ago (Berthelot et al., 2014). Therefore, many genomic regions remain in a pseudo-tetraploid status, which complicates sequence assembly and development of genetic markers because of the difficulty to distinguish true allelic variants from PSVs. Therefore, homozygous individuals were used to produce the first genome sequence and reference transcriptome (Berthelot et al., 2014), subsequent improved genome assemblies (Pearse et al., 2019; Gao et al., 2021), and also to validate the large set of SNPs used in the first 57K SNP chip (Palti et al., 2014, 2015). In the present study, as in Gao et al. (2018), putative PSVs were filtered out by using information from the isogenic lines' genotypes, in order to generate a comprehensive catalogue of reliable SNPs in rainbow trout and then filter out SNPs to be included onto the HD SNP chip.

The 665 K SNP chip was designed based on the Swanson reference genome (Pearse et al., 2019). Only 576K SNPs were uniquely positioned on the Arlee reference genome, which led to a few gaps over 1 Mb based on this reference genome (**Figure 5**) while there was no gap over 250 kb on the Swanson reference genome (**Figure 2**). Genetic and genomic differences between the Swanson and Arlee lines have previously been studied (Palti et al., 2014). It is also known that the two lines differ in their chromosomes' numbers, the Swanson line having $2N$ = 58 with 29 haploid chromosomes (Phillips and Ráb, 2001) and the Arlee line $2N$ = 64 with 32 haploid chromosomes (Ristow et al., 1998). This is not surprising as there are some variable chromosome numbers in rainbow trout populations, associated with Robertsonian centric fusions or fissions, as for instance fission splitting metacentric chromosome 25 observed in Swanson genome into two acrocentric chromosomes in French lines (Guyomard et al., 2012; D'Ambrosio et al., 2019). Depending on the rainbow trout populations, the number of haploid chromosomes (N) varies from 29 to 32 and evidence suggests that the redband trout with $2N = 58$ is the most ancestral type (Thorgaard et al., 1983). In the Arlee karyotype the haploid chromosome number is 32 because chromosomes Omy4, Omy14, and Omy25 are divided into six acrocentric chromosomes (Gao et al., 2021). Note that Arlee chromosomes Omy30, Omy31, and Omy32 correspond to the p-arms of, respectively, Omy4, Omy25, and Omy14 on the Swanson genome.

The 664,531 SNPs successfully genotyped on 463 individuals across two French commercial populations represent a valuable tool for ongoing genomic studies on the genomic architecture of traits, the population evolution history and genetic diversity as well as for the assessment of inbreeding and the genetic effects of management practices in farmed populations. The quality of our new HD chip compared very well with other HD chips developed for salmonid complex genomes. Indeed respectively 90.9% and 91.8% of SNPs were validated in LB and LC samples as high quality based on clustering properties, while corresponding values given for Coho salmon array (Barría et al., 2019) and Atlantic salmon (Yáñez et al., 2016) were 82.5 % and 79.6% for 200 K Affymetrix Axiom® myDesign Custom Arrays. The HD chip is a powerful genomic tool that allow not only to have on average all along the genome a very high density of markers in comparison to the 57 K chip, but also to significantly reduce the number of large gaps (>1 Mb) in the genome coverage. In particular, the extremely low coverage at the telomeric parts of most of the chromosomes or at the centromeric part of metacentric chromosomes have been drastically reduced and the 2 regions spanning over 10 Mb each without any markers on Omy13 (see **Supplementary Figure S1**) have been drastically reduced, leaving just a large gap of 2.75 Mb at the end of Omy13 on the Arlee reference genome. This remaining gap is likely due to the fact that the entire chromosome Omy13 shares high sequence homology with other chromosome arms due to delay in re-diploidization (Gao et al., 2018). The next step will be to develop a new medium-density SNP array for rainbow trout keeping the 39K SNPs present on both the HD chip and the initial 57 K chip, but adding about 25 K SNPs of the HD chip to fill the large gaps without any SNP of the 57 K chip. This second version of the medium-density chip will be a very useful tool both for genomic selection and for cost-effective GWAS thanks to imputation to HD genotypes.

In our study, we illustrate the interest of the HD chip based on LD study across three different rainbow trout populations. The analysis of LD plays a central role in GWAS and fine mapping of QTLs as well as in population genetics to build genetic maps, to estimate recombination rates or effective population sizes as the expected value of $r^2$ is a function of the parameter 4Nec, where c is the recombination rate in Morgan between the markers and $N_e$ is the effective population size (Sved, 1971). The decay and extent of LD at a pairwise distance can be used to determine the evolutionary history of populations (Hayes et al., 2003; Santiago et al., 2020). Lines LB and

LC had the highest LD values in comparison to the American hatchery population (HA), potentially indicating lower effective population sizes in the French selected lines. The lower LD values in the American population may be partly linked to stratification in the sampled population gathered from diverse rivers, but however it helps to quantify the lower bound LD values at short distance that we may expect in hatchery populations. The higher than average LD observed on Omy5 is likely caused by a large chromosomal double-inversion of 55 Mb (Pearse et al., 2019) which prevents recombination in fish.

In rainbow trout farmed populations, the level of strong LD ($r^2 > 0.20$) spans over 100 kb (D'Ambrosio et al., 2019) to 1 Mb (Vallejo et al., 2018; Vallejo et al., 2020). While a number of studies quantify in salmonids the presence of long-range LD from 50 kb to over 1 Mb either for commercial populations (Kijas et al., 2017; Vallejo et al., 2018; Barría et al., 2019; D'Ambrosio et al., 2019) or wild populations (Kijas et al., 2017), little is known on the LD at very short distances. Barría et al. (2019) indicated a maximum value of 0.21 in a Chilean Coho selected line for marker distance lower than 1 kb and a threshold value of *r2* = 0.2 reached at approximately 40 kb. In Atlantic salmon, r2 = 0.2 was reached at approximately 200 kb in a Tasmanian farmed population coming from a single Canadian river without any further introgression (Kijas et al., 2017). In the Tasmanian salmon population, the average LD value for markers separated by 0–10 kb was 0.54 while the corresponding average LD value was only 0.04 in a Finish wild population (Kijas et al., 2017). Previous LD estimates at short distance in French rainbow trout lines ranged from 0.30 to 0.39 at an average distance of 10 kb based on a small set of markers distant from 0 to 20 kb (D'Ambrosio et al., 2019). No LD estimates at inter-marker distances below 10 kb could be reasonably derived in rainbow trout based on the 57 K chip.

In our study, regardless of the rainbow trout populations, the LD values at very short distances between markers (≤10 kb) were moderate (0.44–0.47 at 2 kb and 0.34–0.38 at 10 kb, respectively for LC and LB) compared to the ones observed at similar distances in cattle breeds (Hozé et al., 2013) where r2 values were around 0.70 at 2 kb and in the range 0.50–0.55 at 10 kb whatever the breeds considered. This may be partly due to higher recombination rate in rainbow trout (1.67 cM/Mb; D'Ambrosio et al., 2019) than in cattle (1.25 cM/Mb; Arias et al., 2009), but it also indicates that the founder populations of rainbow trout farmed lines have presumably larger ancestral effective population sizes than cattle breeds. On the contrary, for marker distances over 100 kb, LD values decrease below 0.20 in cattle breeds, while average LD values are still 0.26–0.30 in LC and LB lines, respectively. This indicates stronger recent bottlenecks and selection rates in rainbow trout lines than in cattle breeds. Similar long-range LD was independently observed in two US commercial rainbow trout populations (Vallejo et al., 2018; Vallejo et al., 2020). The pattern of LD decay in rainbow trout commercial lines appears to be more similar to the one observed in conservation flocks of chicken from South Africa (Khanyile et al., 2015), with very similar values reported both at shorter distances than 10 kb, as well as at 500 kb distance where LD values range from 0.15 to 0.24 depending on the conservation flocks and values of 0.21–0.25 were derived for LC and LB, respectively. A last factor that may contribute to this long-range LD in rainbow trout is the high crossing-over interference in males observed when plotting the linkage map distance between markers from the male vs. female linkage maps against the physical distance in base pairs. Sakamoto et al. (2000) have reported a 3.25:1 female to male linkage map distance ratio and Gonzalez-Pena et al. (2016) indicates that female/male recombination ratios were above 2.0 in all the 13 chromosomes known to have homologous pairing with at least one other chromosome arm, while in most of the non-duplicated chromosomes the ratio was generally lower. Because such high crossing-over interference in males were observed in families generated from sex-reversed XX males, we hypothesize that there must be a mechanism that is controlling meiosis in the sperm differently than in the eggs through a different regulation of gene expression not related to presence or absence of the sdY gene.

We have demonstrated in this paper a substantial linkage disequilibrium between neighboring markers, suggesting the density of genotyped SNPs is well-designed to accurately tag most areas of the rainbow trout genome. We acknowledge that, by design, the minor allele frequency distribution of genotyped SNPs is skewed to common alleles, and variation has been predominantly sampled from common SNP shared by both French and North American farmed populations. While this may limit some analyses, we believe that the array will be an invaluable genomic resource for ongoing work investigating genetic diversity, genetic architecture of traits and adaptive potential in world-wide rainbow trout populations.

## DATA AVAILABILITY STATEMENT

Raw sequence data that were generated for French isogenic lines are deposited in the ENA (Projects PRJEB52016 and PRJEB51847). The VCF file for the database of all the SNPs identified in this study including a file with allele frequency information for each SNP is deposited to the public repository ZENODO for downloading under the name "INRAE_USDA_MAF1.vcf.gz" (https://doi.org/10.5281/zenodo.6657091). The sequence and the genotypes of the three French commercial trout lines from "Les Fils de Charles Murgat" (Beaurepaire, France), "Bretagne Truite" (Plouigneau, France) and "Viviers de Sarrance" (Sarrance, France) breeding companies will be made available by request on the recommendation of Pierrick Haffray (SYSAAF, pierrick.haffray@inrae.fr).

## ETHICS STATEMENT

Ethical review and approval was not required for the animal study because this study used fin samples for breeding companies as part of their commercial breeding programs in compliance with Directive 98/58/CE on the protection of animals kept for farming purposes. As such, the project was not subjected to oversight by an institutional

ethic committee. Written informed consent was obtained from the owners for the participation of their animals in this study.

## AUTHOR CONTRIBUTIONS

MB, DL, and FP conceived and designed the study. MD-N, EQ, and PH were involved in the conceptualisation and funding acquisition for the project. YP and GG gave access to the USDA SNP databases. MB, AD, MC, and DL performed bioinformatics analyses on resequencing data. MB, AD, GG, and FP led the design of the 665K SNP array. PP coordinated the collection of samples to be genotyped with the 665K SNP array. JD, LJ, and CP performed the 665K SNP genotyping. KP, HL, MP, and FP analyzed the 665K genotyping data. MB, DL, and FP wrote the manuscript. All authors reviewed and approved the manuscript.

## FUNDING

This study was supported by INRAE, FranceAgrimer and the European Maritime and Fisheries Fund (Hypotemp project, n° P FEA470019FA1000016, and NeoBio project, n° R FEA470016FA1000008). The sequencing of the INRAE rainbow trout isogenic lines were partly funded by CRB-Anim (Biological Resource Centers for Domestic Animals).

## ACKNOWLEDGMENTS

The SNP chip was developed in cooperation with Thermo Fisher and we particularly thank the following Thermo Fisher Scientific personnel for their direct contribution: Ruth Barral Arca, Marie-Laure Schneider, and Philippe Lavisse. We are also grateful to the Genotoul bioinformatics platform (Toulouse Occitanie, doi:10.15454/1.5572369328961167E12) and the *INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi: 10.15454/1.5572390655343293E12)* for providing help, computing, and storage resources. We also thank the 3 French breeding companies "Les fils de Charles Murgat", "Bretagne Truite," and "Viviers de Sarrance" that provided samples for genome resequencing or genotyping on the 665K SNP array.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.941340/full#supplementary-material

## REFERENCES

Altshuler, D. L., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., et al. (2010). A Map of Human Genome Variation from Population-Scale Sequencing. *Nature* 467, 1061–1073. doi:10.1038/nature09534

Anderson, J. L., Rodríguez Marí, A., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P., et al. (2012). Multiple Sex-Associated Regions and a Putative Sex Chromosome in Zebrafish Revealed by RAD Mapping and Population Genomics. *PLoS One* 7, e40701. doi:10.1371/journal.pone.0040701

Arias, J. A., Keehan, M., Fisher, P., Coppieters, W., and Spelman, R. (2009). A High Density Linkage Map of the Bovine Genome. *BMC Genet.* 10, 18. doi:10.1186/1471-2156-10-18

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., et al. (2015). A Global Reference for Human Genetic Variation. *Nature* 526, 68–74. doi:10.1038/nature15393

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One* 3, e3376. doi:10.1371/journal.pone.0003376

Barría, A., Christensen, K. A., Yoshida, G., Jedlicki, A., Leong, J. S., Rondeau, E. B., et al. (2019). Whole Genome Linkage Disequilibrium and Effective Population Size in a Coho Salmon (*Oncorhynchus Kisutch*) Breeding Population Using a High-Density SNP Array. *Front. Genet.* 10, 498. doi:10.3389/fgene.2019.00498

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., et al. (2015). Sex-dependent Dominance at a Single Locus Maintains Variation in Age at Maturity in Salmon. *Nature* 528, 405–408. doi:10.1038/nature16062

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., et al. (2014). The Rainbow Trout Genome Provides Novel Insights into Evolution after Whole-Genome Duplication in Vertebrates. *Nat. Commun.* 5, 3657. doi:10.1038/ncomms4657

Blay, C., Haffray, P., Bugeon, J., D'Ambrosio, J., Dechamp, N., Collewet, G., et al. (2021a). Genetic Parameters and Genome-wide Association Studies of Quality Traits Characterised Using Imaging Technologies in Rainbow Trout, *Oncorhynchus Mykiss*. *Front. Genet.* 12, 639223. doi:10.3389/fgene.2021.639223

Blay, C., Haffray, P., D'Ambrosio, J., Prado, E., Dechamp, N., Nazabal, V., et al. (2021b). Genetic Architecture and Genomic Selection of Fatty Acid Composition Predicted by Raman Spectroscopy in Rainbow Trout. *BMC Genomics* 22, 788. doi:10.1186/s12864-021-08062-7

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Boudry, P., Allal, F., Aslam, M. L., Bargelloni, L., Bean, T. P., Brard-Fudulea, S., et al. (2021). Current Status and Potential of Genomic Selection to Improve Selective Breeding in the Main Aquaculture Species of International Council for the Exploration of the Sea (ICES) Member Countries. *Aquac. Rep.* 20, 100700. doi:10.1016/j.aqrep.2021.100700

Broad Institute (2019). Picard Toolkit: Broad Institute, GitHub Repository. Available at: https://broadinstitute.github.io/picard/ (Accessed December 11, 2017).

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and Applications. *BMC Bioinform.* 10, 421. doi:10.1186/1471-2105-10-421

D'Ambrosio, J., Phocas, F., Haffray, P., Bestin, A., Brard-Fudulea, S., Poncet, C., et al. (2019). Genome-wide Estimates of Genetic Diversity, Inbreeding and Effective Size of Experimental and Commercial Rainbow Trout Lines Undergoing Selective Breeding. *Genet. Sel. Evol.* 51, 26. doi:10.1186/s12711-019-0468-4

D'Ambrosio, J., Morvezen, R., Brard-Fudulea, S., Bestin, A., Acin Perez, A., Guéméné, D., et al. (2020). Genetic Architecture and Genomic Selection of Female Reproduction Traits in Rainbow Trout. *BMC Genomics* 21, 558. doi:10.1186/s12864-020-06955-7

Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., Van Binsbergen, R., Brøndum, R. F., et al. (2014). Whole-genome Sequencing of 234 Bulls Facilitates Mapping of Monogenic and Complex Traits in Cattle. *Nat. Genet.* 46, 858–865. doi:10.1038/ng.3034

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The Variant Call Format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve Years of SAMtools and BCFtools. *Gigascience* 10, 1–4. doi:10.1093/gigascience/giab008

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide Genetic Marker Discovery and Genotyping Using Next-Generation Sequencing. *Nat. Rev. Genet.* 12, 499–510. doi:10.1038/nrg3012

de Roos, A. P. W., Hayes, B. J., Spelman, R. J., and Goddard, M. E. (2008). Linkage Disequilibrium and Persistence of Phase in Holstein-Friesian, Jersey and Angus Cattle. *Genetics* 179, 1503–1512. doi:10.1534/genetics.107.084301

Fairley, S., Lowy-Gallego, E., Perry, E., and Flicek, P. (2020). The International Genome Sample Resource (IGSR) Collection of Open Human Genomic Variation Resources. *Nucleic Acids Res.* 48, D941–D947. doi:10.1093/nar/gkz836

Franěk, R., Baloch, A. R., Kašpar, V., Saito, T., Fujimoto, T., Arai, K., et al. (2020). Isogenic Lines in Fish-A Critical Review. *Rev. Aquacult* 12, 1412–1434. doi:10.1111/raq.12389

Fraslin, C., Brard-Fudulea, S., D'Ambrosio, J., Bestin, A., Charles, M., Haffray, P., et al. (2019). Rainbow Trout Resistance to Bacterial Cold Water Disease: Two New Quantitative Trait Loci Identified after a Natural Disease Outbreak on a French Farm. *Anim. Genet.* 50, 293–297. doi:10.1111/age.12777

Fraslin, C., Phocas, F., Bestin, A., Charles, M., Bernard, M., Krieg, F., et al. (2020). Genetic Determinism of Spontaneous Masculinisation in XX Female Rainbow Trout: New Insights Using Medium Throughput Genotyping and Whole-Genome Sequencing. *Sci. Rep.* 10, 17693. doi:10.1038/s41598-020-74757-8

Gao, G., Nome, T., Pearse, D. E., Moen, T., Naish, K. A., Thorgaard, G. H., et al. (2018). A New Single Nucleotide Polymorphism Database for Rainbow Trout Generated through Whole Genome Resequencing. *Front. Genet.* 9, 147. doi:10.3389/fgene.2018.00147

Gao, G., Magadan, S., Waldbieser, G. C., Youngblood, R. C., Wheeler, P. A., Scheffler, B. E., et al. (2021). A Long Reads-Based De-novo Assembly of the Genome of the Arlee Homozygous Line Reveals Chromosomal Rearrangements in Rainbow Trout. *G3 Genes Genomes Genetics* 11, jkab052. doi:10.1093/g3journal/jkab052

Garrison, E., and Marth, G. (2012). Haplotype-based Variant Detection from Short-Read Sequencing, arXiv [q-bio.GN]. Available at: https://arxiv.org/abs/1207.3907 (Accessed November 13, 2018).

Gautier, M., Laloë, D., and Moazami-Goudarzi, K. (2010). Insights into the Genetic History of French Cattle from Dense SNP Data on 47 Worldwide Breeds. *PLoS One* 5, e13038. doi:10.1371/journal.pone.0013038

Goddard, M. E., Kemper, K. E., MacLeod, I. M., Chamberlain, A. J., and Hayes, B. J. (2016). Genetics of Complex Traits: Prediction of Phenotype, Identification of Causal Polymorphisms and Genetic Architecture. *Proc. R. Soc. B* 283, 20160569. doi:10.1098/rspb.2016.0569

Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B. M., Kenney, P. B., et al. (2016). Genome-Wide Association Study for Identifying Loci that Affect Fillet Yield, Carcass, and Body Weight Traits in Rainbow Trout (*Oncorhynchus Mykiss*). *Front. Genet.* 7, 203. doi:10.3389/fgene.2016.00203

Gorjanc, G., Dumasy, J.-F., Gonen, S., Gaynor, R. C., Antolin, R., and Hickey, J. M. (2017). Potential of Low-Coverage Genotyping-By-Sequencing and Imputation for Cost-Effective Genomic Selection in Biparental Segregating Populations. *Crop Sci.* 57, 1404–1420. doi:10.2135/cropsci2016.08.0675

Gui, J.-F., Zhou, L., and Li, X.-Y. (2022). Rethinking Fish Biology and Biotechnologies in the Challenge Era for Burgeoning Genome Resources and Strengthening Food Security. *Water Biol. Secur.* 1, 100002. doi:10.1016/j.watbs.2021.11.001

Guyomard, R., Boussaha, M., Krieg, F., Hervet, C., and Quillet, E. (2012). A Synthetic Rainbow Trout Linkage Map Provides New Insights into the Salmonid Whole Genome Duplication and the Conservation of Synteny Among Teleosts. *BMC Genet.* 13, 15. doi:10.1186/1471-2156-13-15

Haenel, Q., Laurentino, T. G., Roesti, M., and Berner, D. (2018). Meta-analysis of Chromosome-Scale Crossover Rate Variation in Eukaryotes and its Significance to Evolutionary Genomics. *Mol. Ecol.* 27, 2477–2497. doi:10.1111/mec.14699

Hayes, B. J., and Daetwyler, H. D. (2019). 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annu. Rev. Anim. Biosci.* 7, 89–102. doi:10.1146/annurev-animal-020518-115024

Hayes, B. J., Visscher, P. M., McPartlan, H. C., and Goddard, M. E. (2003). Novel Multilocus Measure of Linkage Disequilibrium to Estimate Past Effective Population Size. *Genome Res.* 13, 635–643. doi:10.1101/gr.387103

Hozé, C., Fouilloux, M.-N., Venot, E., Guillaume, F., Dassonneville, R., Fritz, S., et al. (2013). High-density Marker Imputation Accuracy in Sixteen French Cattle Breeds. *Genet. Sel. Evol.* 45, 33. doi:10.1186/1297-9686-45-33

Johnston, S. E., Bérénos, C., Slate, J., and Pemberton, J. M. (2016). Conserved Genetic Architecture Underlying Individual Recombination Rate Variation in a Wild Population of Soay Sheep (*Ovis Aries*). *Genetics* 203, 583–598. doi:10.1534/genetics.115.185553

Karami, A. M., Ødegård, J., Marana, M. H., Zuo, S., Jaafar, R., Mathiessen, H., et al. (2020). A Major QTL for Resistance to Vibrio Anguillarum in Rainbow Trout. *Front. Genet.* 11, 607558. doi:10.3389/fgene.2020.607558

Kardos, M., Luikart, G., and Allendorf, F. W. (2015). Measuring Individual Inbreeding in the Age of Genomics: Marker-Based Measures Are Better Than Pedigrees. *Heredity* 115, 63–72. doi:10.1038/hdy.2015.17

Khanyile, K. S., Dzomba, E. F., and Muchadeyi, F. C. (2015). Population Genetic Structure, Linkage Disequilibrium and Effective Population Size of Conserved and Extensively Raised Village Chicken Populations of Southern Africa. *Front. Genet.* 6, 13. doi:10.3389/fgene.2015.00013

Kijas, J., Elliot, N., Kube, P., Evans, B., Botwright, N., King, H., et al. (2017). Diversity and Linkage Disequilibrium in Farmed Tasmanian Atlantic Salmon. *Anim. Genet.* 48, 237–241. doi:10.1111/age.12513

Kranis, A., Gheyas, A. A., Boschiero, C., Turner, F., Yu, L., Smith, S., et al. (2013). Development of a High Density 600K SNP Genotyping Array for Chicken. *BMC Genomics* 14, 59. doi:10.1186/1471-2164-14-59

Larson, W. A., Palti, Y., Gao, G., Warheit, K. I., and Seeb, J. E. (2018). Rapid Discovery of SNPs that Differentiate Hatchery Steelhead Trout from ESA-Listed Natural-Origin Steelhead Trout Using a 57K SNP Array. *Can. J. Fish. Aquat. Sci.* 75, 1160–1168. doi:10.1139/cjfas-2017-0116

Li, H. (2013). Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. arXiv [q-bio.GN]. Available at: https://arxiv.org/abs/1303.3997 (Accessed December 11, 2017).

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., et al. (2016). The Atlantic Salmon Genome Provides Insights into Rediploidization. *Nature* 533, 200–205. doi:10.1038/nature17164

Liu, S., Gao, G., Layer, R. M., Thorgaard, G. H., Wiens, G. D., Leeds, T. D., et al. (2021). Identification of High-Confidence Structural Variants in Domesticated Rainbow Trout Using Whole-Genome Sequencing. *Front. Genet.* 12, 639355. doi:10.3389/fgene.2021.639355

Marchini, J., and Howie, B. (2010). Genotype Imputation for Genome-wide Association Studies. *Nat. Rev. Genet.* 11, 499–511. doi:10.1038/nrg2796

Mastrangelo, S., Di Gerlando, R., Di Gerlando, R., Tolone, M., Tortorici, L., Sardina, M. T., et al. (2014). Genome Wide Linkage Disequilibrium and Genetic Structure in Sicilian Dairy Sheep Breeds. *BMC Genet.* 15, 108. doi:10.1186/s12863-014-0108-5

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Meuwissen, T., and Goddard, M. (2010). Accurate Prediction of Genetic Values for Complex Traits by Whole-Genome Resequencing. *Genetics* 185, 623–631. doi:10.1534/genetics.110.116590

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-wide Dense Marker Maps. *Genetics* 157, 1819–1829. doi:10.1093/genetics/157.4.1819

Moen, T., Torgersen, J., Santi, N., Davidson, W. S., Baranski, M., Ødegård, J., et al. (2015). Epithelial Cadherin Determines Resistance to Infectious Pancreatic Necrosis Virus in Atlantic Salmon. *Genetics* 200, 1313–1326. doi:10.1534/genetics.115.175406

Palti, Y., Gao, G., Miller, M. R., Vallejo, R. L., Wheeler, P. A., Quillet, E., et al. (2014). A Resource of Single-Nucleotide Polymorphisms for Rainbow Trout Generated by Restriction-Site Associated DNA Sequencing of Doubled Haploids. *Mol. Ecol. Resour.* 14, 588–596. doi:10.1111/1755-0998.12204

Palti, Y., Gao, G., Liu, S., Kent, M. P., Lien, S., Miller, M. R., et al. (2015). The Development and Characterization of a 57K Single Nucleotide Polymorphism Array for Rainbow Trout. *Mol. Ecol. Resour.* 15, 662–672. doi:10.1111/1755-0998.12337

Pasaniuc, B., Rohland, N., McLaren, P. J., Garimella, K., Zaitlen, N., Li, H., et al. (2012). Extremely Low-Coverage Sequencing and Imputation Increases Power for Genome-wide Association Studies. *Nat. Genet.* 44, 631–635. doi:10.1038/ng.2283

Paul, K., D'Ambrosio, J., and Phocas, F. (2021). Temporal and Region-specific Variations in Genome wide Inbreeding Effects on Female Size and

Reproduction Traits of Rainbow Trout. *Evol. Appl.* 15, 645–662. doi:10.1111/eva.13308

Pearse, D. E., Barson, N. J., Nome, T., Gao, G., Campbell, M. A., Abadía-Cardoso, A., et al. (2019). Sex-dependent Dominance Maintains Migration Supergene in Rainbow Trout. *Nat. Ecol. Evol.* 3, 1731–1742. doi:10.1038/s41559-019-1044-6

Petit, M., Astruc, J.-M., Sarry, J., Drouilhet, L., Fabre, S., Moreno, C. R., et al. (2017). Variation in Recombination Rate and its Genetic Determinism in Sheep Populations. *Genetics* 207, 767–784. doi:10.1534/genetics.117.300123

Phillips, R., and Ráb, P. (2001). Chromosome Evolution in the Salmonidae (Pisces): An Update. *Biol. Rev.* 76, 1–25. doi:10.1017/S1464793100005613

Phocas, F. (2022). "Genotyping, the Usefulness of Imputation to Increase SNP Density, and Imputation Methods and Tools," in *Chapter 4 in: Complex Trait Prediction*. Editors N. Ahmadi and J. Bartholomé (New York, NY: Springer Protocols), 113–138. doi:10.1007/978-1-0716-2205-6_4

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* 81, 559–575. doi:10.1086/519795

Quillet, E., Dorson, M., Leguillou, S., Benmansour, A., and Boudinot, P. (2007). Wide Range of Susceptibility to Rhabdoviruses in Homozygous Clones of Rainbow Trout. *Fish Shellfish Immunol.* 22, 510–519. doi:10.1016/j.fsi.2006.07.002

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Available at: https://www.R-project.org/ (Accessed January 27, 2022).

Reis Neto, R. V., Yoshida, G. M., Lhorente, J. P., and Yáñez, J. M. (2019). Genome-wide Association Analysis for Body Weight Identifies Candidate Genes Related to Development and Metabolism in Rainbow Trout (*Oncorhynchus Mykiss*). *Mol. Genet. Genomics* 294, 563–571. doi:10.1007/s00438-018-1518-2

Ristow, S. S., Grabowski, L. D., Ostberg, C., Robison, B., and Thorgaard, G. H. (1998). Development of Long-Term Cell Lines from Homozygous Clones of Rainbow Trout. *J. Aquatic Animal Health* 10, 75–82. doi:10.1577/1548-8667(1998)010<0075:doltcl>2.0.co;2

Rodríguez, F. H., Flores-Mara, R., Yoshida, G. M., Barría, A., Jedlicki, A. M., Lhorente, J. P., et al. (2019). Genome-wide Association Analysis for Resistance to Infectious Pancreatic Necrosis Virus Identifies Candidate Genes Involved in Viral Replication and Immune Response in Rainbow Trout (*Oncorhynchus Mykiss*). *G3 Genes Genomes Genet.* 9, 2897–2904. doi:10.1534/g3.119.400463

Rodríguez-Ramilo, S. T., Fernández, J., Toro, M. A., Hernández, D., and Villanueva, B. (2015). Genome-Wide Estimates of Coancestry, Inbreeding and Effective Population Size in the Spanish Holstein Population. *PLoS One* 10, e0124157. doi:10.1371/journal.pone.0124157

Sánchez, C. C., Smith, T. P., Wiedmann, R. T., Vallejo, R. L., Salem, M., Yao, J., et al. (2009). Single Nucleotide Polymorphism Discovery in Rainbow Trout by Deep Sequencing of a Reduced Representation Library. *BMC Genomics* 10, 559. doi:10.1186/1471-2164-10-559

Sánchez, C. C., Weber, G. M., Gao, G., Cleveland, B. M., Yao, J., and Rexroad, C. E. (2011). Generation of a Reference Transcriptome for Evaluating Rainbow Trout Responses to Various Stressors. *BMC Genomics* 12, 626. doi:10.1186/1471-2164-12-626

Sakamoto, T., Danzmann, R. G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S. K., et al. (2000). A Microsatellite Linkage Map of Rainbow Trout (*Oncorhynchus Mykiss*) Characterized by Large Sex-specific Differences in Recombination Rates. *Genetics* 155, 1331–1345. doi:10.1093/genetics/155.3.1331

Sandor, C., Li, W., Coppieters, W., Druet, T., Charlier, C., and Georges, M. (2012). Genetic Variants in REC8, RNF212, and PRDM9 Influence Male Recombination in Cattle. *PLoS Genet.* 8, e1002854. doi:10.1371/journal.pgen.1002854

Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., and Caballero, A. (2020). Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Mol. Biol. Evol.* 37, 3642–3653. doi:10.1093/molbev/msaa169

Schweizer, R. M., VonHoldt, B. M., Harrigan, R., Knowles, J. C., Musiani, M., Coltman, D., et al. (2016). Genetic Subdivision and Candidate Genes under Selection in North American Grey Wolves. *Mol. Ecol.* 25, 380–402. doi:10.1111/mec.13364

Silió, L., Rodríguez, M. C., Fernández, A., Barragán, C., Benítez, R., Óvilo, C., et al. (2013). Measuring Inbreeding and Inbreeding Depression on Pig Growth from Pedigree or SNP-Derived Metrics. *J. Anim. Breed. Genet.* 130, 349–360. doi:10.1111/jbg.12031

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., and Smadja, C. M. (2017). Variation in Recombination Frequency and Distribution across Eukaryotes: Patterns and Processes. *Phil. Trans. R. Soc. B* 372, 20160455. doi:10.1098/rstb.2016.0455

Sved, J. A. (1971). Linkage Disequilibrium and Homozygosity of Chromosome Segments in Finite Populations. *Theor. Popul. Biol.* 2, 125–141. doi:10.1016/0040-5809(71)90011-6

Talbert, P. B., and Henikoff, S. (2010). Centromeres Convert but Don't Cross. *PLoS Biol.* 8, e1000326. doi:10.1371/journal.pbio.1000326

Thorgaard, G. H., Allendorf, F. W., and Knudsen, K. L. (1983). Gene-Centromere Mapping in Rainbow Trout: High Interference over Long Map Distances. *Genetics* 103, 771–783. doi:10.1093/genetics/103.4.771

Thorgaard, G. H., Bailey, G. S., Williams, D., Buhler, D. R., Kaattari, S. L., Ristow, S. S., et al. (2002). Status and Opportunities for Genomics Research with Rainbow Trout. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* 133, 609–646. doi:10.1016/s1096-4959(02)00167-7

Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (2017). Genomic Selection Models Double the Accuracy of Predicted Breeding Values for Bacterial Cold Water Disease Resistance Compared to a Traditional Pedigree-Based Model in Rainbow Trout Aquaculture. *Genet. Sel. Evol.* 49, 17. doi:10.1186/s12711-017-0293-6

Vallejo, R. L., Silva, R. M. O., Evenhuis, J. P., Gao, G., Liu, S., Parsons, J. E., et al. (2018). Accurate Genomic Predictions for BCWD Resistance in Rainbow Trout are Achieved Using Low Density SNP Panels: Evidence that Long Range LD is a Major Contributing Factor. *J. Anim. Breed. Genet.* 135, 263–274. doi:10.1111/jbg.12335

Vallejo, R. L., Cheng, H., Fragomeni, B. O., Shewbridge, K. L., Gao, G., MacMillan, J. R., et al. (2019). Genome-wide Association Analysis and Accuracy of Genome-Enabled Breeding Value Predictions for Resistance to Infectious Hematopoietic Necrosis Virus in a Commercial Rainbow Trout Breeding Population. *Genet. Sel. Evol.* 51, 47. doi:10.1186/s12711-019-0489-z

Vallejo, R. L., Fragomeni, B. O., Cheng, H., Gao, G., Long, R. L., Shewbridge, K. L., et al. (2020). Assessing Accuracy of Genomic Predictions for Resistance to Infectious Hematopoietic Necrosis Virus with Progeny Testing of Selection Candidates in a Commercial Rainbow Trout Breeding Population. *Front. Vet. Sci.* 7, 590048. doi:10.3389/fvets.2020.590048

Yáñez, J. M., Naswa, S., López, M. E., Bassini, L., Correa, K., Gilbey, J., et al. (2016). Genomewide Single Nucleotide Polymorphism Discovery in Atlantic Salmon (*Salmo Salar*): Validation in Wild and Farmed American and European Populations. *Mol. Ecol. Resour.* 16, 1002–1011. doi:10.1111/1755-0998.12503

Yoshida, G. M., Carvalheiro, R., Rodríguez, F. H., Lhorente, J. P., and Yáñez, J. M. (2019). Single-step Genomic Evaluation Improves Accuracy of Breeding Value Predictions for Resistance to Infectious Pancreatic Necrosis Virus in Rainbow Trout. *Genomics* 111, 127–132. doi:10.1016/j.ygeno.2018.01.008

Young, W. P., Wheeler, P. A., Fields, A. D., Thorgaard, G. H., and Vrijenhoek, R. C. (1996). DNA Fingerprinting Confirms Isogenicity of Androgenetically Derived Rainbow Trout Lines. *J. Hered.* 87, 77–81. doi:10.1093/oxfordjournals.jhered.a022960

Zeng, Q., Fu, Q., Li, Y., Waldbieser, G., Bosworth, B., Liu, S., et al. (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. doi:10.1038/srep40347

Zhao, Z., Fu, Y.-X., Hewett-Emmett, D., and Boerwinkle, E. (2003). Investigating Single Nucleotide Polymorphism (SNP) Density in the Human Genome and its Implications for Molecular Evolution. *Gene* 312, 207–213. doi:10.1016/S0378-1119(03)00670-X

Zhou, T., Chen, B., Ke, Q., Zhao, J., Pu, F., Wu, Y., et al. (2020). Development and Evaluation of a High-Throughput Single-Nucleotide Polymorphism Array for Large Yellow Croaker (Larimichthys Crocea). *Front. Genet.* 11, 571751. doi:10.3389/fgene.2020.571751

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

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 18 July 2022
doi: 10.3389/fgene.2022.932173

# Development of Microsatellite Markers Based on Transcriptome Sequencing and Evaluation of Genetic Diversity in Swimming Crab (*Portunus trituberculatus*)

*Baohua Duan[1], Shumei Mu[1], Yueqiang Guan[1], Weibiao Liu[1], Tongxu Kang[1], Yana Cheng[1], Zejian Li[2], Yang Tian[3] and Xianjiang Kang[1,4,5]**

[1]College of Life Sciences, Hebei University, Baoding, China, [2]Bureau of Agricultural and Rural Affairs of Huanghua City, Huanghua, China, [3]Hebei Fishery Technology Extension Station, Shijiazhuang, China, [4]Institute of Life Science and Green Development, Hebei University, Baoding, China, [5]Hebei Innovation Center for Bioengineering and Biotechnology, Hebei University, Baoding, China

OPEN ACCESS

**Edited by:**
*Siti Nor, University of Malaysia Terengganu, Malaysia*

**Reviewed by:**
*Xiaoming Pang, Beijing Forestry University, China*
*Dahui Yu, Beibu Gulf University, China*

***Correspondence:**
*Xianjiang Kang*
*xjkang@hbu.edu.cn*

**Specialty section:**
*This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics*

**Received:** *29 April 2022*
**Accepted:** *16 June 2022*
**Published:** *18 July 2022*

**Citation:**
*Duan B, Mu S, Guan Y, Liu W, Kang T, Cheng Y, Li Z, Tian Y and Kang X (2022) Development of Microsatellite Markers Based on Transcriptome Sequencing and Evaluation of Genetic Diversity in Swimming Crab (Portunus trituberculatus). Front. Genet. 13:932173. doi: 10.3389/fgene.2022.932173*

*P. trituberculatus* is an economically important mariculture species in China. Evaluating its genetic diversity and population structure can contribute to the exploration of germplasm resources and promote sustainable aquaculture production. In this study, a total of 246,243 SSRs were generated by transcriptome sequencing of *P. trituberculatus*. Among the examined 254,746 unigenes, 66,331 had more than one SSR. Among the different SSR motif types, dinucleotide repeats (110,758, 44.98%) were the most abundant. In 173 different base repeats, A/T (96.86%), AC/GT (51.46%), and ACC/GGT (26.20%) were dominant in mono-, di-, and trinucleotide, respectively. GO annotations showed 87,079 unigenes in 57 GO terms. Cellular process, cell, and binding were the most abundant terms in biological process, cellular component, and molecular function categories separately. A total of 34,406 annotated unigenes were classified into 26 functional categories according to the functional annotation analysis of KOG, of which "general function prediction only" was the biggest category (6,028 unigenes, 17.52%). KEGG pathway annotations revealed the clustering of 34,715 unigenes into 32 different pathways. Nineteen SSRs were identified as polymorphic and, thus, used to assess the genetic diversity and structure of 240 *P. trituberculatus* individuals from four populations in the Bohai Sea. Genetic parameter analysis showed a similar level of genetic diversity within wild populations, and the cultured population indicated a reduction in genetic diversity compared with wild populations. The pairwise $F_{ST}$ values were between 0.001 and 0.04 with an average of 0.0205 ($p < 0.05$), suggesting a low but significant level of genetic differentiation among the four populations. Structure analysis demonstrated that the four populations were classified into two groups including the cultured group and other populations. The phylogenetic tree and PCA revealed that a vast number of samples were clustered together and that cultivated individuals were distributed more centrally than wild individuals. The findings contribute to

the further assessment of germplasm resources and assist to provide valuable SSRs for marker-assisted breeding of *P. trituberculatus* in the future.

**Keywords: *Portunus trituberculatus*, transcriptome, SSR, genetic diversity, Bohai Sea**

## INTRODUCTION

Swimming crab, *P. trituberculatus*, is an edible portunid of great commercial significance, which has been widely farmed in China. Due to its high nutritional value and rapid growth, *P. trituberculatus* has become one of the most important economic crab species in marine aquaculture (Lv et al., 2017). Indeed, *P. trituberculatus* is one of the most heavily fished brachyurans in the world with approximately 95% of the total catch occurring in China (Liu et al., 2013; Hui et al., 2018). The total catch was 424,630 tons in 2020 (Bureau of Fisheries of Ministry of Agriculture, PRC, 2021). With increasing serious problems such as mass outbreaks of disease, overfishing, and water pollution in recent years, however, the *P. trituberculatus* aquaculture and fishing industry face great pressure. Despite the improved artificial propagation and rearing techniques, the crab industry still relies on the collection of wild specimens to provide parental stock (Wang et al., 2012). It is urgent to determine the genetic diversity and population structure of *P. trituberculatus* for the exploration of germplasm resources and conservation management of this species.

Genetic diversity is the material basis necessary for populations to deal with changing environment, and it can trace the history of biological evolution and explore the evolutionary potential of existing organisms (Liu et al., 2021a). However, for the breeding of shrimp and crab, long-term artificial directional selection eventually leads to a decline in genetic diversity in the population (Wang et al., 2018a). Moreover, it is difficult for natural stocks to recover from declining genetic diversity caused by overfishing (Liu et al., 2009). Benefitting from the rapid advance of high-throughput sequencing and genotyping technologies, an increasing number of molecular markers are developed and applied to genetic analyses in aquatic species. To date, molecular markers including isozyme (Fan et al., 2009), random amplified polymorphic DNA (RAPD) (Chi et al., 2010), amplified fragments length polymorphism (AFLP) (Liu et al., 2013; Liu et al., 2014), mitochondrial DNA (Guo et al., 2012; Shan et al., 2017; Hui et al., 2018), microsatellites DNA (Lee et al., 2013; Yue et al., 2022), and single nucleotide polymorphism (SNP) (Duan et al., 2022a) were developed and used in population genetic analysis of *P. trituberculatus*. Among these markers, microsatellite DNA markers (simple sequence repeats, SSRs) have become an ideal molecular marker in population genetics research because of their co-dominant inheritance, high polymorphism, reproducibility, hyper-variable, transferability, random distribution in the genome, and ease of analysis *via* PCR (Gou et al., 2020; Pavan Kumara et al., 2020; Zhu et al., 2021; Lu et al., 2022). Such markers are often used to obtain genetic diversity coefficients, which can provide a basis for genetic protection strategies. However, the traditional methods of developing SSRs are usually time-consuming and labor-intensive because of establishing of the genomic library to get the fragmented sequence and hybridization *in situ* with probes. In recent years, the increased access and affordability of high-throughput sequencing technologies have enabled genomic and transcriptomic research on many marine species, thus leading to more rapid and accurate identification of SSR markers (Yang et al., 2018; Liu et al., 2022).

In recent years, transcriptome sequencing (RNA-seq) is widely used in the study of species genetics because of its wide dynamic range, precise, sensitivity, unbiased quantification of transcripts, and comprehensive coverage of all expressed sequences in a given tissue sample (Chakraborty et al., 2022). Now, RNA-seq is a very updated and efficient method for discovering new genes, expression pattern identification, and development of SSR markers with higher throughput and much lower cost (Li et al., 2018; Tulsania et al., 2020). SSR markers acquired from RNA-seq have intrinsic advantages over genomic SSRs because of high efficiency, strong transferability, and correlation with potential genes, as well as their applicability as anchor markers for comparative mapping and evolutionary studies (Zheng et al., 2014; Zhao et al., 2019). Transcriptomic SSRs have been extensively explored and applicated in various aquatic species, such as giant freshwater prawn (*Macrobrachium rosenbergii*) (Jung et al., 2011; Yu et al., 2019), oriental river prawn (*Macrobrachium nipponense*) (Ma et al., 2012), mud crab (*Scylla paramamosain*) (Ma et al., 2014), *Paphia textile* (Chen et al., 2016), *Penaeus monodon* (Nguyen et al., 2016), and pearl oyster (*Pinctada maxima*) (Wang et al., 2019a). Transcriptomics have also played an important role in advancing portunid aquaculture (Waiho et al., 2022). Nevertheless, the current transcriptome studies involving *P. trituberculatus* mainly focus on its nutrition (Zhou et al., 2019a; Fang et al., 2021a), development (Liu et al., 2018a; Liu et al., 2019), reproduction (Wang et al., 2018b), and sex determination (Zhang et al., 2022), while molecular research is scarce.

The present study aimed to 1) develop SSR markers with RNA-seq technology; 2) characterize the transcriptome of *P. trituberculatus*; and 3) evaluate the genetic diversity and structure among different populations using the polymorphic SSR markers from transcriptome sequencing. Our findings not only contribute to molecular genetic analyses but also provide valuable information for effective breeding and conservation strategies of the *P. trituberculatus* aquaculture.

## MATERIALS AND METHODS

### Sample Collection and DNA Extraction

A total of 240 swimming crab samples from four populations were collected (**Supplementary Table S1**; **Figure 1**), including

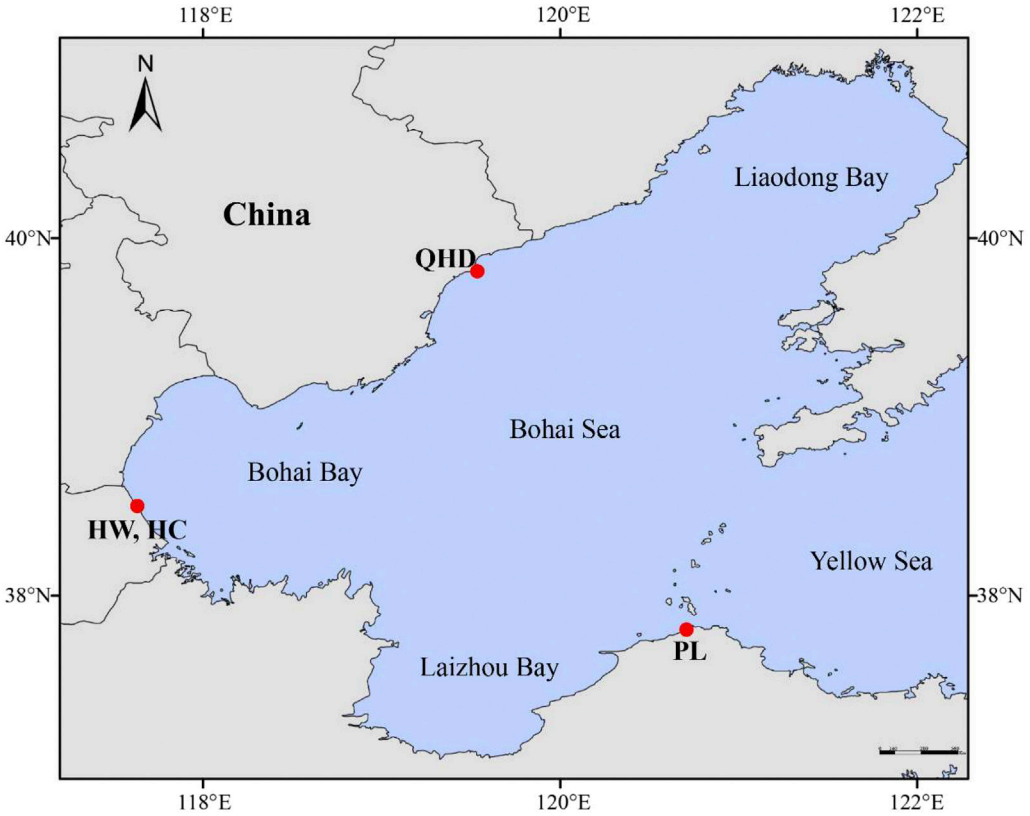



**FIGURE 1 |** Collection map of *P. trituberculatus* samples.

three wild populations [Qinhuangdao (QHD), Huanghua (HW), and Penglai (PL)] from the Bohai Sea and one cultured stock (HC) from the national breeding farm of swimming crabs in Huanghua (Hebei, China) which is adjacent to the Bohai Sea. Claws of each sample were obtained and preserved in absolute alcohol and stored at −20°C until DNA extraction. Genomic DNA was isolated from the claw muscle using the TIANamp Marine animal DNA extraction kit (TIANGEN, Beijing, China) following the manufacturer's protocols. The quality and concentration of the extracted DNA were determined using the NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, DE, United States) and 1% agarose electrophoresis gel, and then diluted to 100 ng/μl and stored at −20°C for polymerase chain reaction (PCR) amplification.

## RNA Extraction, Library Preparation, and Transcriptome Sequencing

A total of 18 individuals *P. trituberculatus* (nine wild females and nine cultured females) were collected from the national breeding farm of swimming crabs in Huanghua, China. The crabs were all anesthetized on ice and dissected to collect muscle and ovaries samples. All of the samples were rapidly flash-frozen in liquid nitrogen and stored at −80°C for RNA extraction. An equal amount of either muscle tissue or ovaries tissue was dissected from each of three individuals and pooled into one mixed sample. A total of 12 mixed samples were produced, including 6 mixed muscle samples (3 wild and 3 farmed) and 6 mixed ovarian samples (3 wild and 3 farmed), each containing three biological replicates. Total RNA was extracted from each mixed sample using the TRIzol reagent (Invitrogen, Carlsbad, CA, United States) following the manufacturer's protocol. RNA degradation and contamination were monitored on 1% agarose gels, and RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, United States). RNA concentration and integrity were measured using the Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, United States) and the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, United States), respectively.

A total amount of 1.5 μg RNA from each sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) following the manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer (5X). First-strand cDNA was synthesized using random hexamer primer and M-MuLV

Reverse Transcriptase (RNase $H^-$). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. Remaining overhangs were converted into blunt ends *via* exonuclease/polymerase activities. After adenylation of 3′ ends of DNA fragments, NEBNext Adaptor with hairpin loop structure was ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 250–300 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, United States). Then 3 μl USER Enzyme (NEB, United States) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95°C before PCR. Subsequently, PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers, and Index (X) Primer. Then PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. At last, the library preparations were sequenced on an Illumina HiSeq 4000 platform and paired-end reads were generated at Novogene Corporation (Tianjin, China).

## Quality Control, Transcriptome Assembly, and Gene Function Annotation

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. In this step, the clean data (clean reads) were obtained by removing reads containing adapter, reads containing ploy-N, and low-quality reads (quality score < 20) from raw data. At the same time, Q20, Q30, GC-content, and sequence duplication levels of the clean data were calculated. All the downstream analyses were based on clean data with high quality. Transcriptome assembly of the high-quality clean reads was accomplished using Trinity software with default settings (Grabherr et al., 2011). In order to annotate the assembled unigenes, a BLASTX search with an E-value < $10^{-5}$ (Camacho et al., 2009) was performed against several public databases, including NCBI non-redundant protein sequences (Nr), NCBI non-redundant nucleotide sequences (Nt), Protein family (PFAM), euKaryotic Ortholog Groups (KOG), Swiss-Prot protein, KEGG Ortholog database (KO) and Gene Ontology (GO). Assigning the GO terms to the unigenes was implemented on Blast2GO software (Götz et al., 2008).

## Simple Sequence Repeats Identification and Primer Design

SSR loci were identified throughout all unigenes generated by the *P. trituberculatus* transcriptome sequencing using MISA software version 1.0 (http://pgrc.ipk-gatersleben.de/misa/misa.html). The minimum number of repeats was defined as ten for mononucleotide repeats, six for dinucleotide repeats, five for tri-, tetra-, penta-, and hexanucleotide repeats. Primer pairs for each SSR locus were designed using Primer3 (http://primer3.sourceforge.net/releases.php) according to the following criteria: 1) primer length of 18–25 bp; 2) annealing temperature (Tm) between 55°C and 62°C; 3) GC content from 40% to 60%; 4) PCR product length of 90–250 bp; 5) avoidance of primer dimers and hairpin structures. SSR primers were synthesized by General Biosystems Co., Ltd. (Anhui, China). Twelve samples were used to identify the polymorphism of the selected SSR primers through PCR amplification and 8% non-denaturing polyacrylamide gel electrophoresis with pBR322 DNA/MspI (MBI) as a standard DNA marker.

**TABLE 1 |** Characteristics of 19 SSR loci for *P. trituberculatus*.

| Locus | Primer sequences | Repeat motif | Anneal (°C) |
|---|---|---|---|
| PrMa01 | F:CCTTGCCTCGTCAGTGTCAT<br>R:TGGCTGTAGACACCCTCCAT | (CTG)6 | 60 |
| PrMa02 | F:AGAGCTGACCTCGCTTTGAC<br>R:TCCAGCTCCTCCTGTCCAAT | (GTG)8 | 60 |
| PrMa03 | F:CTTGATTGCCTCTCGCTTGT<br>R:GGGGGAGAGGGAGAGAATGT | (TG)10 | 60 |
| PrMa04 | F:TCCTGGACCTTGTTCAGTCC<br>R:GCAATCCCACACACACTCCT | (TCC)10 | 60 |
| PrMa05 | F:GCGTTGCGTGTACTGAAAGT<br>R:GCGGCTCTGGTCAGGAATAC | (TG)31 | 60 |
| PrMa06 | F:TCCTGCAACTTACATTCTTGGTC<br>R:GTGTGCACAGGATACAGCCT | (CA)15 | 60 |
| ZL05 | F:AGAATGTTGCCATGGCTGGA<br>R:ACCCTGTATCAGTGCGTTGG | (GGT)7 | 60 |
| ZL06 | F:CCCGCCCCTGTACATTTTCA<br>R:TGTTGGTAGGCTTGGTGGTC | (TAT)10 | 60 |
| ZL08 | F:GCTTCTGCTGCTGGTCCTTA<br>R:ACCAGACATTGCTGAGCATG | (CAAC)10 | 60 |
| DX05 | F:GTGGGCCGCCAATATCACTA<br>R:AATCCACCACTTGCACCCAA | (TG)12 | 60 |
| DX07 | F:CGTGCATCCGTGTGTTTGTT<br>R:GCCATCTTTTCGCCGAGTTG | (TG)10 | 60 |
| DX09 | F:TAGGCATGGGATGGGTGAGA<br>R:CGGGAAGGAGTGTTGTTGAGT | (CA)17 | 60 |
| DX10 | F:AATCACAACCCAGCCGCATA<br>R:ACAACGAAGGAGAGATGCGG | (TG)12 | 60 |
| DX14 | F:CCCGCTACCCCATAACTCAC<br>R:TCTTCCTCCCCACAGCCATA | (GTG)7 | 60 |
| DX15 | F:CGTCCCATCATCTGACAAAGG<br>R:TCCTTCACCTCTTCCTCTTTTCT | (GAG)6 | 60 |
| DX16 | F:GAGGCAAGCAAGTTAACCATTAG<br>R:CTTCCTGGTTACCTCATCCTACC | (GT)7 | 60 |
| DX19 | F:CACACTCGTTGCAGACACTACTT<br>R:CTGTTACTTACTCGGTGCTTTGG | (TG)11 | 60 |
| TRAN2 | F:TCACTACCACTACCGCTTTGTTT<br>R:GATGTCAGTAACGGGAGAGTGAG | (CAC)8 | 60 |
| TRAN3 | F:GCTGTTGTAGAAACCCATGAAAG<br>R:AGGGAGATACACGACCAACACTA | (GTG)7 | 60 |

## Simple Sequence Repeat Genotyping

A total of 19 pairs of polymorphic SSR primers were identified and used for subsequent analysis (**Table 1**). All forward primers were labeled with the fluorescent dye, 6-carboxy-fluorescein (FAM). PCR amplification was performed in 20 μl reaction volumes containing 2 μl of template DNA, 2 μl of each primer (2.5 μmol/L each), 10 μl of 2 × Es Taq Master Mix (CWBIO, Beijing, China), and 4 μl of $ddH_2O$. Amplification cycles consisted of initial denaturation (5 min at 95°C), followed by 34 cycles of denaturation (30 s at 95°C), annealing (30 s at 60°C), extension (30 s at 72°C), and further extension (10 min at 72°C). After amplification, PCR products were diluted 10 times in sterile water. The pooled sample was composed of 20 μl Hi-Di formamide and 0.2 μl GeneScan 500 ROX Size Standard. An

**TABLE 2 |** Summary statistics for transcriptome sequencing of *P. trituberculatus*.

| Category | Number |
|---|---|
| Total raw reads | 661,922,456 |
| Total clean reads | 637,983,466 |
| Clean reads proportion (%), Q20 (%), Q30 (%) | 99.38, 96.3, 91.07 |
| Total number of unigenes examined | 254,746 |
| Mean length of unigenes (bp), N50 (bp) | 1,077, 1,936 |
| GC content (%) | 50.88 |
| Total amount of transcripts | 338,285 |
| Mean length of transcripts (bp), N50 (bp) | 879, 1,730 |
| Total size of examined sequences (bp) | 274,270,543 |
| Total number of identified SSRs | 246,243 |
| Number of SSR containing sequences | 132,908 |
| Number of sequences containing more than one SSR | 66,331 |

ABI 3730XL Genetic Analyzer (Applied Biosystems, Foster City, CA, United States) was used to conduct capillary electrophoresis (CE) following the manufacturer's instructions. Each CE sample contained 1 μl diluted PCR product and 15 μl pooled sample. Allele sizes (in base pairs) were determined with GeneMarker® Fragment Analysis Software (Softgenetics LLC®, State College, PA, United States) on the comparison of the position of the internal size standard in each lane with the position of the peak value of each sample.

## Genetic Diversity and Population Structure Analysis

Genetic diversity was estimated by determining the genetic parameters, such as the number of alleles (*Na*), the effective number of alleles (*Ne*), the observed heterozygosity ($H_O$) and expected heterozygosity ($H_E$) using POPGENE version 1.3 (Yeh et al., 1999). Based on allele frequency, the polymorphism information content (*PIC*) per locus was estimated by PIC-CALC software (Nagy et al., 2012). The null allele frequency (*F*na) for loci was calculated using GenePOP (Rousset, 2008). *p* values were calculated for determining the Hardy–Weinberg equilibrium (HWE) at each locus with POPGENE version 1.3. Genetic differentiation index ($F_{ST}$) between populations and analysis of molecular variance (AMOVA) were calculated using GenAlEx 6.5 (Peakall and Smouse, 2012).

The phylogenetic tree was constructed using the neighbor-joining (NJ) method as implemented in MEGA7 (Kumar et al., 2016). Principal component analysis (PCA) was carried out using Canoco 4.5 to elucidate genetic relationships within and among populations. Bayesian model-based population genetic structure was inferred by STRUCTURE version 2.3.4 (Pritchard et al., 2000). The putative number of populations (K) was set from 1 to 10 with 3 replicate simulations for each K value using 100,000 MCMC (Markov Chain Monte Carlo) iterations after an initial 100,000 burn-in period. With the log probability of data (LnP(D)) and an *ad hoc* statistic ΔK based on the rate of change in LnP(D) between successive K-values, the structure output was entered into Structure Harvester (Evanno et al., 2005; Earl and Vonholdt, 2012) to determine the optimum K value.

# RESULTS

## Transcriptome Assembly and Sequence Annotation

The transcriptome sequencing of 12 mixed samples from muscle and ovary of *P. trituberculatus* was conducted to generate RNA sequences, and the statistical data has been shown in **Table 2**. Illumina sequencing generated 661,922,456 raw reads. The raw reads produced in this study have been deposited in the Short Read Archive of the National Center for Biotechnology Information (NCBI) with accession numbers SUB11453401 and PRJNA836158. After stringent quality filtering, a total of 637,983,466 clean reads were obtained, accounting for 99.38% of the total raw reads. GC content ranged from 48.5% to 53.47% with an average of 50.88%, and the mean Q20 and Q30 were 96.3% and 91.07%, respectively. A total of 338,285 transcripts were identified with an average length of 879 bp (N50 length of transcript = 1,730 bp, which is defined as the shortest sequence length of 50% of total contigs and is used to evaluate the quality of assembled sequences), of which 11,0596 (32.69%) were less than 301 bp in length; 84,886 (25.09%) were 301–500 bp; 61,915 (18.3%) were 501–1,000 bp; 43,085 (12.74%) were 1,001–2,000 bp; 37,803 (11.18%) were over 2,000 bp (**Supplementary Figure S1**). Totaling 254,746 unigenes were assembled with an average length of 1,077 bp (N50 length of unigenes is 1,936 bp), among which 47,174 (18.52%) were less than 301 bp in length; 67,149 (26.36%) were 301–500 bp; 59,692 (23.43%) were 501–1,000 bp; 42,928 (16.85%) were 1,001–2,000 bp; 37,803 (14.84%) were over 2,000 bp (**Supplementary Figure S1**).

GO database was the largest matched database with 87,079 unigenes (34.18% of all unigenes) annotated, followed by PFAM (86,669, 34.02%), Nr (77,856, 30.56%), SwissProt (58,305, 22.88%), KO (34,715, 13.62%), KOG (34,406, 13.5%), and Nt (29,269, 11.48%) database (**Supplementary Figure S2**). In all, 118,572 (46.54%) unigenes were annotated in at least one database and 9,901 (3.88%) were annotated in all databases. In the Nr databases, 77,856 unigenes were annotated from 835 species. The top-hit species in similarity search against the Nr database included *Zootermopsis nevadensis* (9,854, 12.7%), *Daphnia pulex* (4,798, 6.2%), *Tribolium castaneum* (2,730, 3.5%), *Stegodyphus mimosarum* (2,399, 3.1%), *Crassostrea gigas* (2,389, 3.1%), and other (55,686, 71.5%) (**Supplementary Figure S3**).

For the functional annotation and classification of the assembled unigenes, 87,079 unigenes were assigned to 57 GO terms which included three ontology categories: biological process (258,416 unigenes), cellular component (178,933), and molecular function (102,370) (**Figure 2A**). The main components within biological process category contained cellular process (52,412, 60.19%), metabolic process (43,020, 49.40%), and single-organism process (39,395, 45.24%). Cell (32,536, 37.36%) and cell part (32,535, 37.36%) were the most frequent proportion in cellular component category. In the molecular function category, the largest potion was binding (45,970, 52.79%), followed by catalytic activity (32,373, 37.18%). According to KOG annotations, 34,406 annotated unigenes

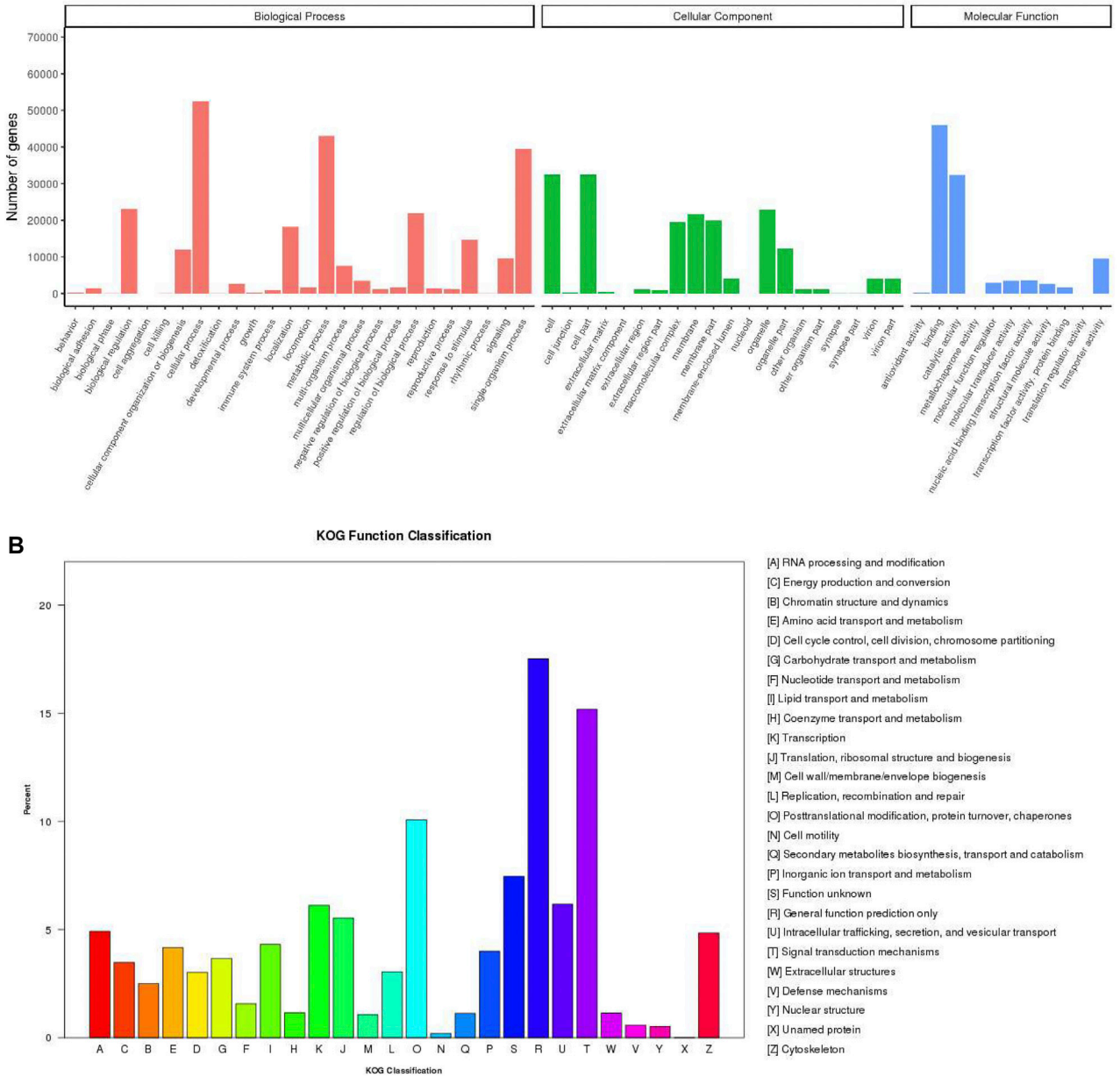



**FIGURE 2 |** Annotation of the *P. trituberculatus* transcriptome. **(A)** GO annotation and classification of unigenes. **(B)** KOG annotation and classification of unigenes. The x-axis indicates the categories, and the y-axis indicates the number of the unigenes.

were classified into 26 functional categories (**Figure 2B**). Among these categories, "general function prediction only" was the biggest category (6,028 unigenes, 17.52%), followed by "signal transduction mechanisms" (5,223, 15.18%), and "posttranslational modification, protein turnover, chaperones" (3,466, 10.02%) category.

Identification of the biological pathways was performed according to the KEGG annotations, which showed the clustering of 34,715 unigenes into 32 pathways (**Figure 3**). Detailedly, these unigenes were categorized into five KEGG biochemical pathways: Cellular Processes (A), Environmental Information Processing (B), Genetic Information Processing (C), Metabolism (D), and Organismal Systems (E). This analysis revealed that the top five pathways included signal transduction (4,665 unigenes, 13.44%), endocrine system (2,629, 7.57%), transport and catabolism (2,491, 7.18%), translation (2,100, 6.05%), and cellular community (1,852, 5.33%).

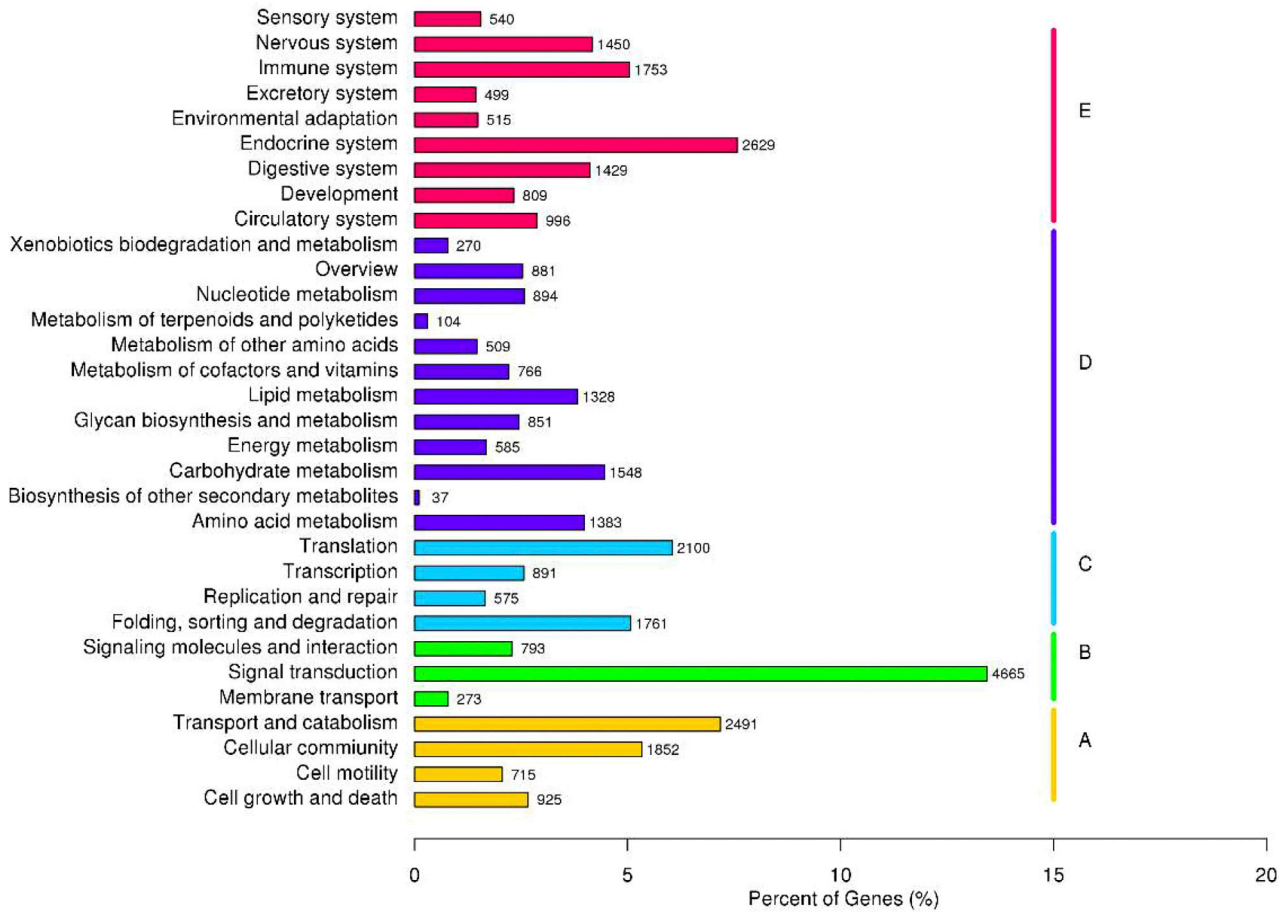



**FIGURE 3 |** Function annotation and classification of *P. trituberculatus* unigenes in the KEGG category.

## Characterization of Simple Sequence Repeat Markers

A total of 246,243 SSRs were identified from 132,908 SSR-containing unigenes of *P. trituberculatus*. The distribution density was one SSR per 1.11 kb, and 66,331 sequences contained more than one SSR (**Table 2**). Dinucleotide repeats were the most abundant type (110,758, 44.98%), followed by mono- (64,679, 26.27%), tri- (61,923, 25.15%), tetra- (7,154, 2.91%), penta- (1,343, 0.55%), and hexanucleotide (386, 0.16%) (**Supplementary Table S1**). The number of tandem repeats of microsatellite motifs ranged from 5 to 103. Microsatellites with six tandem repeats (28371, 11.52%) were the most common, followed by five tandem repeats (25,920, 10.53%), ten tandem repeats (24,777, 10.06%), and eleven tandem repeats (24,136, 9.88%). Microsatellite motifs with > 16 tandem repeats accounted for 22.39% (55,129) (**Supplementary Table S2**).

Of the 246,243 SSR loci, 173 different repeat motifs were detected (**Supplementary Table S3**). Among the two types of mononucleotide repeats, A/T was the most abundant (62,646, 96.85%) when compared to C/G. AC/GT (56,994, 51.46%) was the dominant motif type in the dinucleotide repeat, followed by AG/CT (47,487, 42.87%), and AT/AT (6,046, 5.46%). The most abundant types in the trinucleotide were ACC/GGT (16,224, 26.20%). Moreover, a low percentage (3.62%) of tetra-, penta- and hexanucleotide repeat motifs were observed in all identified microsatellite motifs. The physical positions of these SSR markers in the unigenes were also identified that 1708, 8,644, and 5,067 SSRs were located in the coding sequence (CDS), 3′untranslated region (UTR) and 5′UTR, respectively. In CDS, trinucleotide repeats (1,279, 74.88%) were the dominant type. Most of the mono- and dinucleotide repeats (6,290, 72.77%) were located in 3′UTR, and 5′UTR contained the majority of di- and trinucleotide repeats (3,283, 64.79%) (**Supplementary Figure S4**).

A total of 104,424 pairs of SSR primers were designed, 150 of which were randomly selected to identify polymorphism (**Supplementary Table S4**). Finally, 19 pairs of SSR primers

**TABLE 3 |** Genetic diversity parameters for 19 SSR loci.

| Locus | *Na* | *Ne* | $H_O$ | $H_E$ | *PIC* | *Fna* | HWE |
|---|---|---|---|---|---|---|---|
| PrMa01 | 12 | 5.124 | 0.925 | 0.805 | 0.779 | 0.0000 | ** |
| PrMa02 | 11 | 7.354 | 0.532 | 0.864 | 0.849 | 0.0706 | ** |
| PrMa03 | 21 | 11.161 | 0.833 | 0.91 | 0.903 | 0.0486 | ** |
| PrMa04 | 12 | 6.672 | 0.808 | 0.85 | 0.834 | 0.0157 | ** |
| PrMa05 | 30 | 20.156 | 0.791 | 0.95 | 0.948 | 0.084 | ** |
| PrMa06 | 20 | 9.693 | 0.488 | 0.897 | 0.888 | 0.2086 | ** |
| ZL05 | 9 | 4.239 | 0.467 | 0.764 | 0.727 | 0.0991 | ** |
| ZL06 | 11 | 5.079 | 0.708 | 0.803 | 0.782 | 0.0788 | ** |
| ZL08 | 12 | 6.644 | 0.82 | 0.85 | 0.833 | 0.000 | ns |
| DX05 | 10 | 6.363 | 0.871 | 0.843 | 0.824 | 0.000 | ns |
| DX07 | 7 | 3.892 | 0.613 | 0.743 | 0.702 | 0.0254 | ** |
| DX09 | 21 | 12.708 | 0.774 | 0.921 | 0.916 | 0.0835 | ** |
| DX10 | 10 | 4.938 | 0.699 | 0.798 | 0.773 | 0.0525 | ** |
| DX14 | 8 | 4.052 | 0.525 | 0.753 | 0.714 | 0.0289 | ** |
| DX15 | 13 | 8.75 | 0.571 | 0.886 | 0.875 | 0.1829 | ** |
| DX16 | 7 | 4.222 | 0.679 | 0.763 | 0.727 | 0.0359 | ** |
| DX19 | 11 | 6.647 | 0.938 | 0.85 | 0.832 | 0.000 | ** |
| TRAN2 | 9 | 4.367 | 0.629 | 0.771 | 0.736 | 0.000 | ** |
| TRAN3 | 9 | 2.751 | 0.483 | 0.636 | 0.598 | 0.000 | ** |
| Mean | 12.79 | 7.095 | 0.692 | 0.824 | 0.802 | | |

Na, *number of alleles;* Ne, *number of effective alleles;* Ho, *observed heterozygosity;* $H_E$, *expected heterozygosity;* PIC, *polymorphism information content;* Fna, *frequency of null alleles; HWE, Hardy–Weinberg equilibrium;* **p < *0.01. ns, no deviations from HWE.*

**TABLE 4 |** Mean genetic parameters of four *P. trituberculatus* populations.

| Population | *Na* | *Ne* | $H_O$ | $H_E$ | *PIC* |
|---|---|---|---|---|---|
| HW | 11.421 | 5.572 | 0.706 | 0.756 | 0.728 |
| PL | 11.842 | 5.595 | 0.7 | 0.76 | 0.731 |
| QHD | 11.263 | 5.802 | 0.675 | 0.752 | 0.724 |
| HC | 8.737 | 4.507 | 0.688 | 0.716 | 0.679 |

showed high polymorphism in 8% non-denaturing polyacrylamide gel electrophoresis (**Table 1**), and they were used for subsequent analysis in *P. trituberculatus*.

## Genetic Diversity Within Populations

All parameters of the 19 SSR loci were calculated (**Table 3**). The *PIC* values ranged from 0.598 (TRNA3) to 0.948 (PrMa05) with a mean of 0.802, showing that these SSRs has high polymorphism (*PIC* > 0.5) and are suitable for the evaluation of genetic diversity in *P. trituberculatus* populations. A total of 243 alleles were found with an average of 12.79 per locus. The *Ne* values ranged from 2.751 to 20.156 with a mean of 7.095. The $H_O$ and $H_E$ ranged from 0.467 to 0.938 (mean: 0.692) and from 0.636 to 0.95 (mean: 0.824), respectively. The three wild populations (HW, PL, and QHD) showed a similar level of genetic diversity, while the cultivated population indicated a reduction in genetic diversity compared with them due to the relatively smaller genetic parameters (**Table 4**). Notably, the majority of SSR loci had null alleles and presented significant deviations from HWE, and heterozygote deficiency ($H_O < H_E$) was observed, with the exception of PrMa01, DX05, and DX19.

**TABLE 5 |** Estimates of pairwise $F_{ST}$ values among the four *P. trituberculatus* populations.

| | HW | HC | PL | QHD |
|---|---|---|---|---|
| HW | — | | | |
| HC | 0.040* | — | | |
| PL | 0.016* | 0.023* | — | |
| QHD | 0.020* | 0.023* | 0.001* | — |

**Significant difference (*p < *0.05).*

## Population Differentiation and Variation

The populations HW and HC showed highest $F_{ST}$ (0.040) ($p < 0.05$) whereas the lowest $F_{ST}$ of 0.001 ($p < 0.05$) was observed between populations PL and QHD (**Table 5**). The mean $F_{ST}$ was observed to be 0.0205, indicating low but significant levels of genetic differentiation among the four populations. The results of AMOVA revealed that only 2% of genetic variation was partitioned among populations while 98% of the variation was concentrated within populations (**Table 6**).

## Population Genetic Structure

The genetic structural analysis of 240 *P. trituberculatus* samples was performed to infer the optimal K value with the ΔK method. When the highest ΔK value was observed, the optimal K value was 2 (**Figure 4A**), indicating that all individuals were clustered into two groups, including wild group (green) and the cultivated group (red) (**Figure 4B**). This was consistent with the population-level phylogenetic tree that the four populations were divided into 2 main clusters, in which cluster 1 contained only HC and cluster 2 contained all wild populations (**Supplementary Figure S5**). A certain degree of biological mixing, however, was also observed between wild and cultivated samples. The PCA revealed that the first two principal components (PCs) explained 7.9% (PC1) and 6.62% (PC2) of total variation respectively (**Figure 5**). The majority of samples were clustered together, and no obvious geographical patterns were observed. The cultivated individuals were mainly clustered towards the right side (positive values) of PC1. The individual-level phylogenetic tree was constructed based on NJ method, in which all individuals were clustered into two clades, and no significant clustering patterns related to geographical locations were found, but cultivated individuals were distributed more centrally than wild individuals (**Figure 6**).

# DISCUSSION

## Transcriptome Sequencing

*P. trituberculatus* is a typical mariculture species with significant economic value. The Bohai Sea is an important habitat and fishing area for this species. The germplasm resources from the Bohai Sea form a vital foundation for the protection and breeding of *P. trituberculatus* in the national breeding farms of swimming crabs. Large-scale development of molecular markers and advancement of high-throughput sequencing technologies provide a solid support for germplasm resources assessment of *P. trituberculatus* in recent years (Shan et al., 2017; Liu et al., 2021b).

**TABLE 6 |** Analysis of molecular variance (AMOVA) from four *P. trituberculatus* populations.

| Source of variation | *df* | *SS* | *MS* | Variance component | Percentage of variation (%) | *p*-value |
|---|---|---|---|---|---|---|
| Among populations | 3 | 77.042 | 25.681 | 0.149 | 2 | 0.001 |
| Within populations | 476 | 3408.842 | 14.333 | 7.166 | 98 | 0.001 |
| Total | 479 | 3485.883 | | 7.316 | 100 | 0.001 |

df, *degrees of freedom;* SS, *sum of squares;* MS, *mean square.*

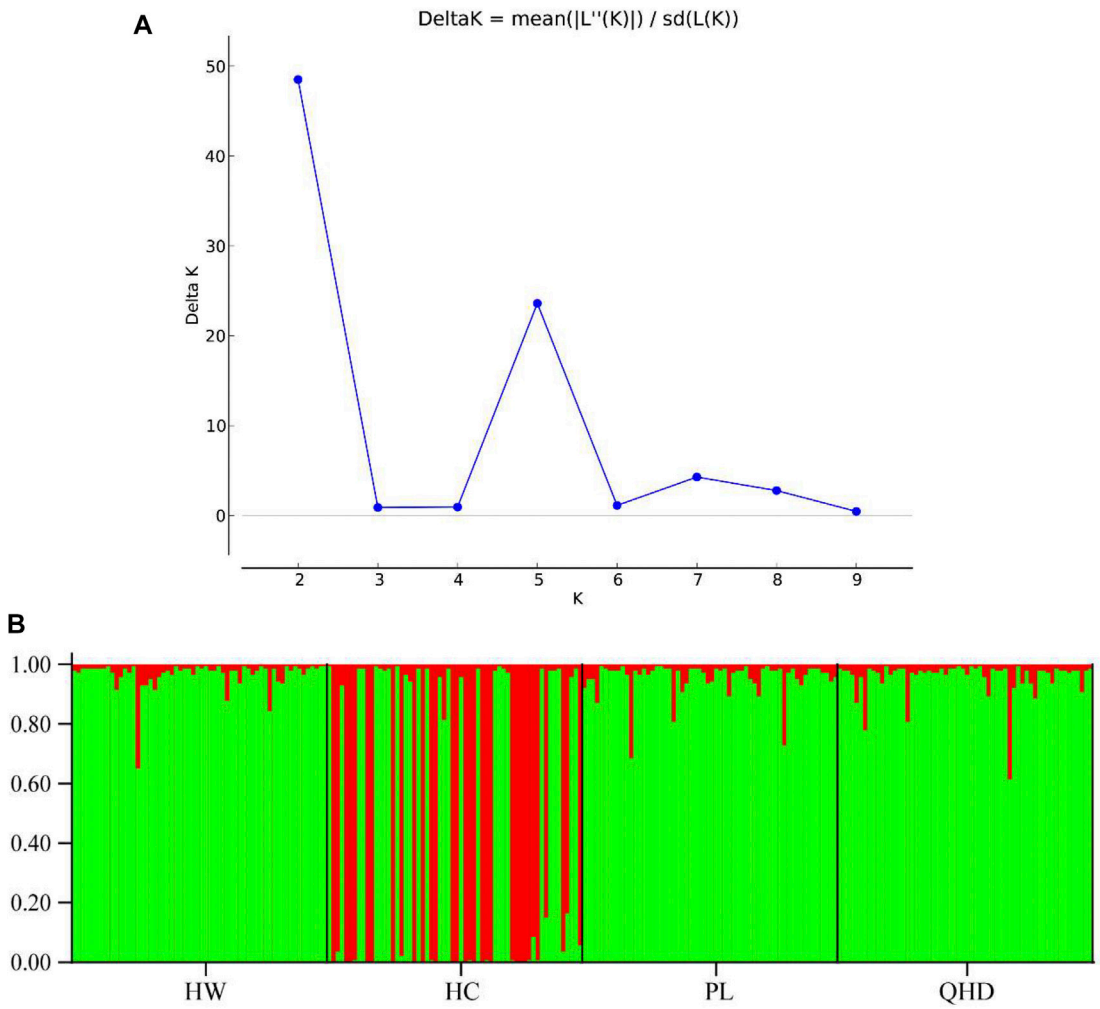



**FIGURE 4 | (A)** Relationships between the number of clusters (K) and the corresponding. ΔK statistics calculated based on STRUCTURE analysis. **(B)** Population genetic structure based on the Bayesian clustering model among 240 *P. trituberculatus* samples at K = 2.

Lv et al. (2014) identified 22,673 SSR markers of *P. trituberculatus* with transcriptome sequencing, which provided a material basis for future genetic linkage and quantitative trait loci analyses in this species. In this study, Illumina sequencing of *P. trituberculatus* was performed for the development of SSR markers, which generated new high-throughput data for transcriptomics, providing valuable information for germplasm conservation and selective breeding of *P. trituberculatus*.

A total of 246,243 SSRs were identified with the Illumina HiSeq 4000 platform, which was far higher than that of yesso scallop (*Patinopecten yessoensis*) (2,748 SSRs, Hou et al., 2011), yellow drum (*Nibea albiflora*) (12,254, Gong et al., 2016), Xingguo red carp (*Cyprinus carpio* var.*singuonensis*) (13,652, Yue et al., 2016), freshwater ornamental shrimps (*Neocaridina denticulate*) (25,355, Huang et al., 2020), and clam (*Cyclina sinensis*) (12,418, Fang et al., 2020). The possible reason is that

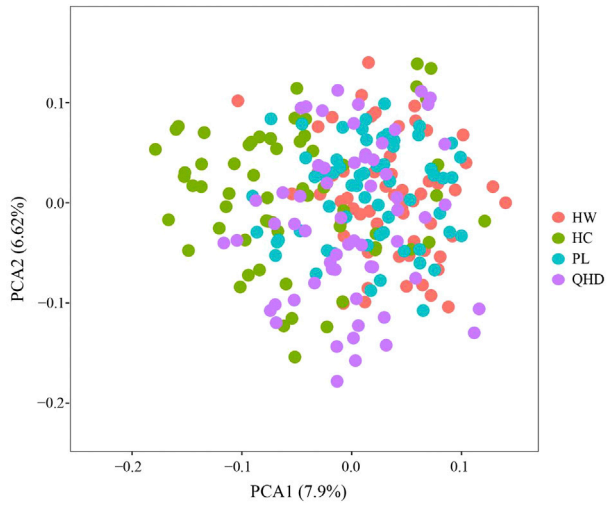



**FIGURE 5 |** PCA analyses based on SSR data for 240 individuals from 4 *P. trituberculatus* populations.

*P. trituberculatus* has more chromosomes (2n = 106) than shellfish, fish, and other crustaceans (Liu et al., 2012), thus containing more SSR sequences. The average density of SSRs was 1/1.11 kb, which was higher than that in *P. maxima* (1/29.73 kb) (Wang et al., 2019a), sea cucumber (*Apostichopus japonicus*) (1/29.2 kb) (Du et al., 2012), mandarin fish (*Siniperca chuatsi*) (1/26.28 kb) (Sun et al., 2019), large-scale loach (*Paramisgurnus dabryanus*) (1/6.99 kb) (Li et al., 2015a), giant wrasse (*Cheilinus undulatus*) (1/5.35 kb) (Liu et al., 2020), burbot (*Lota lota*) (1/4.25 kb) (Meng et al., 2019), and the Bombay duck (*Harpadon nehereus*) (1/3 kb) (Huang et al., 2021), while lower than that of the South African abalone (*Haliotis midae*) (1/0.756 kb) (Franchini et al., 2011), tu-chung (*Eucommia ulmoides*) (1/0.73 kb) (Huang et al., 2013). The differences between SSR frequency and density may be attributed to genome structure, SSR mining tools, dataset size, and search criteria (Liu et al., 2021c). In addition, the most abundant types of SSR are dinucleotide repeats, which was consistent with the conclusions obtained by other high-throughput sequencing technologies in aquatic animals (Kong et al., 2019; Zhai et al., 2020).

In 173 different repeat motifs from the identified microsatellites in this study, A/T was the most abundant motif type higher than C/G in mononucleotide, which was congruent with the previous studies (Song et al., 2008; Fang et al., 2020; Tian et al., 2021). Ni et al. (2018) indicated that DNA recombination and replication sliding mechanisms in PCR amplification might result in high A/T content. In addition, methylated cytosine C is easily mutated into thymine T through deamination, which makes G/C mutate A/T in the process of DNA replication and transcription (Schlötterer and Tautz, 1992). In dinucleotide repeats, AC/GT exhibited the highest frequency, which was coincident with the result obtained by Lv et al. (2013). In the development of SSR markers based on transcriptome sequencing in blood clam (*Scapharca kagoshimensis*), AC/GT was also predominant among dinucleotide repeats (Chen et al., 2022). However, the opposite results were found in the Zhikong scallop (*Chlamys farreri*) (Zhan et al., 2008) and bay scallop (*Argopecten irradians*) (Li et al., 2009) from the SSR-enriched library, which showed that the proportion of AG/CT was higher than that of AC/GT in the genome. This difference may be related to the SSR screening method, base composition preference of different coding genes, and methylase activity *in vivo* (Chen et al., 2022).

The location of SSR loci determines their functional roles. SSRs in CDS affect the inactivated or activated genes or protein synthesis process, and SSRs in 3′UTR are involved in transcription slippage or gene silencing, and SSRs in 5′UTR impact gene transcription and translation (Xia et al., 2014; Xu et al., 2020; Liu et al., 2021c). In the present study, 88.92% of microsatellites were located in UTRs, which was much higher than that of CDS regions. One possible reason is that microsatellites located in UTRs are subject to fewer evolutionary constraints and natural selection pressure, thus easily leading to phenotype changes (Xu et al., 2020; Vidya et al., 2021). Moreover, 74.88% of trinucleotide repeats were found to be accumulated in CDSs regions. This might explain that non-trinucleotides negatively selected frameshift mutations, while trinucleotide did not cause frame shift mutation and failed to affect gene expression (Xia et al., 2014; Liu et al., 2021c).

To explore the potential functions of the obtained unigenes, the functional annotation and classification of these unigenes were conducted through BLASTX search in the public databases. GO annotations showed a lot of unigenes distributed to cellular process, metabolic process, cell, cell part, binding, and catalytic activity terms. This suggests that genes encoding these functions may be more conserved across different species and are thus easier to annotate in the database. In addition, KEGG and KOG annotations revealed that many unigenes might participate in the life activities and basal metabolism of *P. trituberculatus* with various biological functions. In summary, these annotation analyses contribute to finding potential genes associated with the growth and development of *P. trituberculatus* for breeding programs. Further studies also should be carried out to identify the molecular functions of these putative genes.

## Population Genetic Diversity

Genetic diversity is the foundational core of ecosystems and species diversity and can reveal population connectivity and adaptive potential of a species as well as provide insight into past events (Fang et al., 2021b; Ma et al., 2021). It is affected by many factors, including artificial selection, genetic drift, migration, and breeding systems and is usually evaluated by the polymorphism information content (*PIC*) and heterozygosity (Zhou, et al., 2019b). Values of *PIC* above 0.5 indicate high polymorphism (Singh et al., 2020). Heterozygosity is an important index to evaluate population variation at the genetic level, and the greater its value, the higher the population genetic diversity (Qin et al., 2013). Li et al. (2011) investigated the genetic diversity of five *P. trituberculatus* populations with eight SSR markers and observed that the mean $H_E$ values ranged from 0.7283 to 0.7704, which revealed a high level of genetic diversity in the

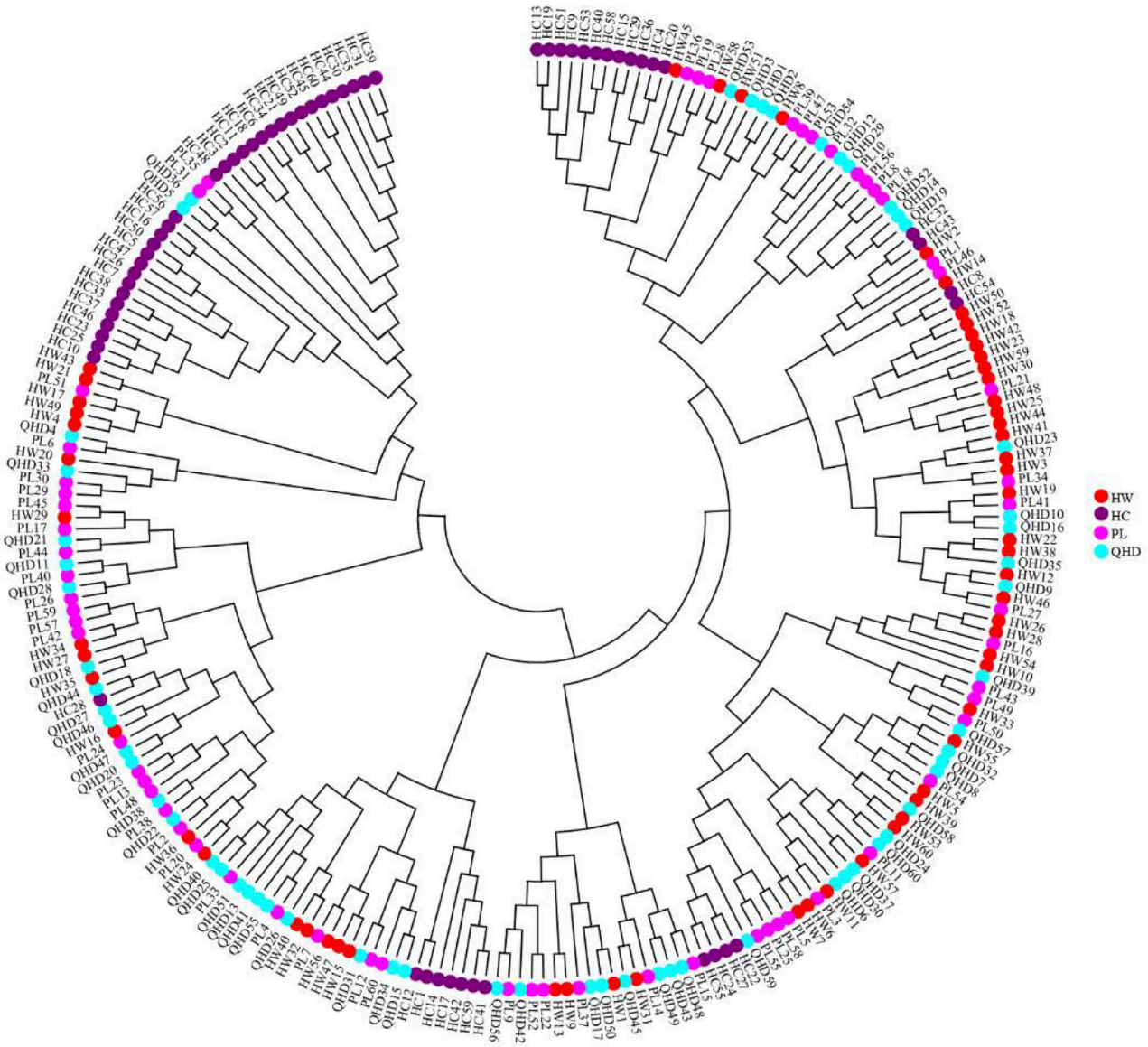



**FIGURE 6 |** Neighbor-joining phylogenetic tree of 240 individuals of *P. trituberculatus* based on SSR data.

wild resources. The current study reports *PIC* values of all SSR loci greater than 0.5, indicating the high polymorphic nature of the loci and their suitability for assessing genetic diversity in the four *P. trituberculatus* populations. The observed and expected heterozygosity values indicated a similar level of genetic diversity among the wild populations, and compared with the wild populations, the genetic diversity of the cultivated population showed a reduction because of lower genetic coefficients. In the estimation of genetic diversity of *P. trituberculatus* populations from Shandong peninsula, Liu et al. (2012) found a similar result. It is possible that genetic decline, genetic drift, and inbreeding result in low genetic variability in farmed stocks (Jorge et al., 2018). Additionally, the domesticated stocks are subjected to artificial selection in a selective breeding program, which may show reduced effective population size, thus leading to a decline in genetic diversity (Wong et al., 2022).

In general, the expected heterozygosity ($H_E$) is more accurate than the observed heterozygosity ($H_O$) for evaluating the level of population genetic diversity because *Ho* is easily influenced by sample sizes (Qin et al., 2013). Based on polymorphic SSR markers, middle ($H_E$ = 0.73–0.84) to high ($H_E$ = 0.916–0.918) genetic diversity of *P. trituberculatus* was revealed by Guo et al. (2013) and Xu and Liu (2011). In this study, the mean $H_E$ ranged from 0.675 to 0.706 which is lower than those observed in the above studies, thus showing a lower level of genetic diversity among the four populations. This may be attributable to the special geographical location of the Bohai Sea. The Bohai Sea is a semi-enclosed shallow water body and has limited connectivity to

the Yellow Sea by the Bohai Strait, which restricts the dispersal of crabs and thus results in low genetic diversity (Liu et al., 2012). Moreover, the phenomena of eutrophication and hypoxia, as well as the serious interference from anthropogenic activities such as land-source pollution, aquaculture pollution, and reclamation in the Bohai Sea also reduced the genetic diversity of species (Wang et al., 2021).

## Genetic Differentiation and Variation Among Populations

$F_{ST}$ is an important gauge of genetic differentiation between populations and is crucial for a better understanding of the genetic relationships. A value of $F_{ST}$ which falls between 0 and 0.05 shows a low level of genetic differentiation (Wang et al., 2019b). The current study reports the mean $F_{ST}$ values of 0.021 ($p < 0.05$) which is less than 0.05 (**Table 5**), indicating low levels of genetic differentiation among the four *P. trituberculatus* populations, which is conformed to the result described by Xu and Liu (2011). Based on the pairwise $F_{ST}$ ranging from 0.0142 to 0.0498, six SSR loci showed no genetic difference between wild and cultivated populations of *P. trituberculatus* from the Zhejiang coastal region (Li et al., 2015b). This genetic similarity may be accounted for release of hatchery-produced offspring, which results in hybrid germplasm. Liu et al. (2018b) and Liu et al. (2021b) used SSR markers to evaluate the population structure of *P. trituberculatus* in Panjin and Yingkou (Liaoning, China) adjacent to Liaodong Bay, respectively, and found low but significant levels of genetic differentiation ($F_{ST} < 0.05$, $p < 0.05$), suggesting that large-scale stock enhancement of *P. trituberculatus* presents potential genetic risks on wild populations, and that the relevant management measures should be formulated to achieve successful stock enhancement and resource restoration for the swimming crab.

During the current study, AMOVA results revealed that total variance within populations (98%) was significantly greater than that among populations (2%). The result corresponded to the genetic variation found in blue swimmer crab (*Portunus pelagicus*) (Chai et al., 2017). Most loci showed a deficit of heterozygotes, which might result from the presence of null alleles, artificial selection, migration, and inbreeding in the population (Guo et al., 2013). Additionally, a majority of SSR loci deviated from the Hardy-Weinberg equilibrium, and this finding might be ascribed to null alleles or a small number of samples. Hence, designing more effective SSR primers to eliminate null alleles, and combining more molecular markers with a larger sample size are essential to elaborate the genetic diversity of *P. trituberculatus* populations in the Bohai Sea. In addition, genetic monitoring is required to preserve the genetic variations for preventing germplasm degradation and making full use of the genetic resources of *P. trituberculatus*.

## Population Genetic Structure

A stable genetic structure is central to species survival. Its disintegration will lead to decreased populations or even extinction. Given the economic significance of *P. trituberculatus*, the understanding genetic structure is crucial for the development of effective management strategies and can provide a genetic tool for breeding and offer a scientific support for resource conservation of this species (Liu et al., 2009). The results of the current study establish that K = 2 is the most likely number of clusters when ΔK is at its highest. This finding confirms that the *P. trituberculatus* specimens from the four populations cluster into two groups including the cultured group and the wild group (**Figure 4**). Some of the genetic information gained from the cultured samples has been assigned to wild populations, indicating that the ancestral generation of these wild individuals may derive from cultivated populations because of the hatchery-reared seed release activities. In a cultivated group, some genetic information that derives from wild samples can be observed. This observation indicates some degree of introgression of wild populations into the cultivated population, which may be accounted for the fact that fertilized female crabs are caught as broodstocks from the wild to use to artificially culture and produce seeds (Duan et al., 2022b).

The individual-level phylogenetic tree and PCA illuminated that all individuals showed some degree of genetic connectivity, and that the cultured individuals were relatively concentrated in comparison with wild individuals. Despite the annual release activities, the gene exchange between cultivated and wild populations is limited when compared to that between wild populations in the open sea, thus leading to the separation of cultured individuals from all individuals, showing more obvious particularity. Therefore, it is vital to further investigate the genetic structure of wild and cultured populations of *P. trituberculatus* in the Bohai Sea for formulating scientific management measures to prevent mutual interference between them.

# CONCLUSION

Overall, this study performed assembly of transcriptome sequences, functional annotation, and SSR markers discovery of *P. trituberculatus*. Nineteen polymorphic SSRs were identified and used to investigate the genetic variation and structure of the four *P. trituberculatus* populations from the Bohai Sea. The findings revealed a lower level of genetic diversity in *P. trituberculatus* populations from the Bohai Sea when compared to the other populations from the Yellow Sea and the East China Sea. The pairwise $F_{ST}$ values showed low but significant genetic differentiation between populations. The population structure analysis, phylogenetic tree, and PCA showed a mixing of wild and cultivated individuals, which corroborated the genetic connectivity between them, but cultivated individuals were distributed more centrally than wild individuals. In addition, heterozygote deficiencies, null alleles, and significant deviation from HWE at many SSR loci were observed. Therefore, practical and effective measures are expected to be taken to reinforce the identification and protection of genetic diversity and prevent degeneration of *P. trituberculatus* germplasm. For example, developing high-quality markers such as SNPs using a chromosome-level genome of *P. trituberculatus* (Tang et al., 2020; Lv et al., 2021), and carrying out a large-scale investigation to fully elucidate the genetic diversity and

population structure of *P. trituberculatus* in the Bohai Sea. Additionally, increasing the scale of swimming crab aquaculture, extending the fishing moratorium, and performing long-term genetic monitoring is also helpful for the conservation and utilization of germplasm resources in *P. trituberculatus*. In conclusion, the results improve our understanding of the population genetic structure of *P. trituberculatus* in the Bohai Sea and provide valuable information for the selection breeding of this species.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found in NCBI under accession numbers SUB11453401 and PRJNA836158.

## AUTHOR CONTRIBUTIONS

BD: writing—original draft, formal analysis, investigation, and methodology. SM and YG: data curation and formal analysis. WL, TK, and YC: investigation, methodology, and validation. ZL and YT: resources. XK: supervision, project administration, conceptualization, and writing—review and editing. All authors contributed to the article and approved the submitted version.

## FUNDING

This research was supported by the Natural Science Foundation of Hebei Province (C2016201249); Science and Technology Innovation Project of Modern Seed Industry (21326307D); Institute of Life Science and Green Development (Hebei University), and Innovation Center for Bioengineering and Biotechnology of Hebei Province.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.932173/full#supplementary-material

## REFERENCES

Bureau of Fisheries of Ministry of Agriculture, PRC (2021). *China Fishery Statistical Yearbook of 2021*. Beijing, China: Agricultural press.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: Architecture and Applications. *BMC Bioinforma.* 10, 421. doi:10.1186/1471-2105-10-421

Chai, C. J., Bin Esa, Y., Ismail, S., and Kamarudin, M. S. (2017). Population Structure of the Blue Swimmer Crab *Portunus Pelagicus* in Coastal Areas of Malaysia Inferred from Microsatellites. *Zool. Stud.* 56, e26. doi:10.6620/ZS.2017.56-26

Chakraborty, D., Sharma, N., Kour, S., Sodhi, S. S., Gupta, M. K., Lee, S. J., et al. (2022). Applications of Omics Technology for Livestock Selection and Improvement. *Front. Genet.* 13, 774113. doi:10.3389/fgene.2022.774113

Chen, L., Li, L., Shi, X., Qin, Y., Liu, L., and Guo, Y. (2022). Development and Evaluation of SSR Markers Based on Transcriptome Sequencing in Scapharca Kagoshimensis. *Prog. Fish. Sci.* 43, 129–137. doi:10.19663/j.issn2095-9869.20210206003

Chen, X., Li, J., Xiao, S., and Liu, X. (2016). De Novo assembly and Characterization of Foot Transcriptome and Microsatellite Marker Development for Paphia Textile. *Gene* 576, 537–543. doi:10.1016/j.gene.2015.11.001

Chi, D., Yan, B., Shen, S., and Gao, H. (2010). Rapd Analysis between Color-Different Crab Individuals of Portunus Trituberculatus. *Mar. Sci. Bull.* 12, 47–54. doi:10.3969/j.issn.1000-9620.2010.02.006

Du, H., Bao, Z., Hou, R., Wang, S., Su, H., Yan, J., et al. (2012). Transcriptome Sequencing and Characterization for the Sea Cucumber *Apostichopus Japonicus* (Selenka, 1867). *PLoS One* 7, e33311. doi:10.1371/journal.pone.0033311

Duan, B., Mu, S., Guan, Y., Li, S., Yu, Y., Liu, W., et al. (2022a). Genetic Diversity and Population Structure of the Swimming Crab (*Portunus Trituberculatus*) in China Seas Determined by Genotyping-By-Sequencing (GBS). *Aquaculture* 555, 738233. doi:10.1016/j.aquaculture.2022.738233

Duan, B., Liu, W., Li, S., Yu, Y., Guan, Y., Mu, S., et al. (2022b). Microsatellite Analysis of Genetic Diversity in Wild and Cultivated *Portunus Trituberculatus* in Bohai Bay. *Mol. Biol. Rep.* 49, 2543–2551. doi:10.1007/s11033-021-07054-w

Earl, D. A., and Vonholdt, B. M. (2012). STRUCTURE HARVESTER: A Website and Program for Visualizing Structure Output and Implementing the Evanno Method. *Conserv. Genet. Resour.* 4, 359–361. doi:10.1007/s12686-011-9548-7

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the Number of Clusters of Individuals Using the Software Structure: A Simulation Study. *Mol. Ecol.* 14, 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x

Fan, X., Gao, B., Liu, P., and Li, J. (2009). Genetic Variation Analysis of Four Wild Populations of *Portunus Trituberculatus* by Isozyme. *Prog. Fish. Sci.* 30, 84–89. doi:10.3969/j.issn.1000-7075.2009.04.013

Fang, F., Yuan, Y., Jin, M., Shi, B., Zhu, T., Luo, J., et al. (2021a). Hepatopancreas Transcriptome Analysis Reveals the Molecular Responses to Different Dietary N-3 PUFA Lipid Sources in the Swimming Crab Portunus Trituberculatus. *Aquaculture* 543, 737016. doi:10.1016/j.aquaculture.2021.737016

Fang, J., Shen, Y., Zhang, L., Li, T., and Zhao, Y. (2020). Development and Verification of SSR Markers in Cyclina Sinensis. *J. Phys. Oceanogr.* 39, 214–220. doi:10.3969/J.ISSN.2095-4972.2020.02.008

Fang, Y., Chen, J., Ruan, H., Xu, N., Que, Z., and Liu, H. (2021b). Genetic Diversity and Population Structure of *Metaphire Vulgaris* Based on the Mitochondrial Coi Gene and Microsatellites. *Front. Genet.* 12, 686246. doi:10.3389/fgene.2021.686246

Franchini, P., Van der Merwe, M., and Roodt-Wilding, R. (2011). Transcriptome Characterization of the South African Abalone *Haliotis Midae* Using Sequencing-By-Synthesis. *BMC Res. Notes* 4, 59. doi:10.1186/1756-0500-4-59

Gong, S., Wang, Z., Xiao, S., Lin, A., and Xie, Y. (2016). Development and Verification of SSR Based on Transcriptome of Yellow Drum. *Nibea Albiflora. J. Jimei Univ. Nat. Sci.* 21, 241–246. doi:10.19715/j.jmuzr.2016.04.001

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-Throughput Functional Annotation and Data Mining with the Blast2Go Suite. *Nucleic Acids Res.* 36, 3420–3435. doi:10.1093/nar/gkn176

Gou, X., Shi, H., Yu, S., Wang, Z., Li, C., Liu, S., et al. (2020). SSRMMD: A Rapid and Accurate Algorithm for Mining SSR Feature Loci and Candidate Polymorphic SSRs Based on Assembled Sequences. *Front. Genet.* 11, 706. doi:10.3389/fgene.2020.00706

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. *Nat. Biotechnol.* 29, 644–652. doi:10.1038/nbt.1883

Guo, E., Cui, Z., Wu, D., Hui, M., Liu, Y., and Wang, H. (2013). Genetic Structure and Diversity of *Portunus Trituberculatus* in Chinese Population Revealed by Microsatellite Markers. *Biochem. Syst. Ecol.* 50, 313–321. doi:10.1016/j.bse.2013.05.006

Guo, E., Liu, Y., Cui, Z., Li, X., Cheng, Y., and Wu, X. (2012). Genetic Variation and Population Structure of Swimming Crab (*Portunus Trituberculatus*) Inferred

from Mitochondrial Control Region. *Mol. Biol. Rep.* 39, 1453–1463. doi:10.1007/s11033-011-0882-3

Hou, R., Bao, Z., Wang, S., Su, H., Li, Y., Du, H., et al. (2011). Transcriptome Sequencing and De Novo Analysis for Yesso Scallop (Patinopecten Yessoensis) Using 454 GS FLX. *PLoS One* 6, e21560. doi:10.1371/journal.pone.0021560

Huang, C.-W., Chu, P.-Y., Wu, Y.-F., Chan, W.-R., and Wang, Y.-H. (2020). Identification of Functional SSR Markers in Freshwater Ornamental Shrimps *Neocaridina Denticulata* Using Transcriptome Sequencing. *Mar. Biotechnol.* 22, 772–785. doi:10.1007/s10126-020-09979-y

Huang, H., Du, H., Wuyun, T., and Liu, P. (2013). Development of SSR Molecular Markers Based on Transcriptome Sequencing of. *Eucommia Ulmoides. Sci. Silvae Sin.* 49, 176–181. doi:10.11707/j.1001-7488.20130523

Huang, X., Jiang, Y., Jiang, X., and Yang, T. (2021). Analysis of Microsatellite Markers in *Harpadon Nehereus* Based on Transcriptome Sequencing Illumina Hiseq$^{tm}$ 2500. *J. Zhejiang Ocean. Univ. Nat. Sci.)* 40, 189–197. doi:10.3969/j.issn.1008-830X.2021.03.001

Hui, M., Shi, G., Sha, Z., Liu, Y., and Cui, Z. (2018). Genetic Population Structure in the Swimming Crab, *Portunus Trituberculatus* and its Implications for Fishery Management. *J. Mar. Biol. Ass.* 99, 891–899. doi:10.1017/S0025315418000796

Jorge, P. H., Mastrochirico-Filho, V. A., Hata, M. E., Mendes, N. J., Ariede, R. B., Freitas, M. V. d., et al. (2018). Genetic Characterization of the Fish *Piaractus Brachypomus* by Microsatellites Derived from Transcriptome Sequencing. *Front. Genet.* 9, 46. doi:10.3389/fgene.2018.00046

Jung, H., Lyons, R. E., Dinh, H., Hurwood, D. A., McWilliam, S., Mather, P. B., et al. (2011). Transcriptomics of a Giant Freshwater Prawn (*Macrobrachium Rosenbergii*): De Novo Assembly, Annotation and Marker Discovery. *PLoS One* 6, e27938. doi:10.1371/journal.pone.0027938

Kong, X., Li, M., Chen, Z., Gong, Y., Zhang, J., and Zhang, P. (2019). Development and Evaluation of Di-/Tri-Nucleotide-repeated Microsatellites by RAD-Seq in. *Decapterus Macrosoma. South China Fish. Sci.* 15, 97–103. doi:10.12131/20180256

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054

Lee, H. J., Lee, D.-H., Yoon, S.-J., Kim, D. H., Kim, S.-G., Hyun, Y. S., et al. (2013). Characterization of 20 Microsatellite Loci by Multiplex PCR in Swimming Crab, Portunus Trituberculatus. *Genes. Genom* 35, 77–85. doi:10.1007/s13258-013-0062-z

Li, C., Ling, Q., Ge, C., Ye, Z., and Han, X. (2015a). Transcriptome Characterization and SSR Discovery in Large-Scale Loach *Paramisgurnus dabryanus* (Cobitidae, Cypriniformes). *Gene* 557, 201–208. doi:10.1016/j.gene.2014.12.034

Li, H., Liu, X., Du, X., Song, R., Zhang, G., Hu, J., et al. (2009). Development of Microsatellite Markers in Bay Scallop and Their Inheritance Patterns in an F1 Hybrid Family. *Mar. Sci.* 33, 4–8.

Li, N., Zheng, Y.-Q., Ding, H.-M., Li, H.-P., Peng, H.-Z., Jiang, B., et al. (2018). Development and Validation of SSR Markers Based on Transcriptome Sequencing of *Casuarina equisetifolia. Trees* 32, 41–49. doi:10.1007/s00468-017-1607-6

Li, P., Xu, K., and Zhou, H. (2015b). Genetic Diversity of *Portunus Trituberculatus* Among the Cultured and Wild Populations in the Offing of Zhejiang Using Microsatellite Markers. *J. Zhejiang Ocean. Univ. Nat. Sci.)* 34, 537–542. doi:10.3969/j.issn.1008-830X.2015.06.007

Li, X., Li, P., Li, J., and Gao, B. (2011). Genetic Diversity Among Five Wild Populations of Portunus Trituberculatus. *J. Fish. Sci. China* 18, 1327–1334. doi:10.3724/SP.J.1118.2011.01327

Liu, B., Zhang, X., Wang, Z., Li, W., Zhang, Q., Liu, Q., et al. (2021b). Genetic Pattern Fluctuations in Wild Swimming Crab Populations, Under the Influence of Continuous Mass Stock Enhancement. *Fish. Res.* 243, 106075. doi:10.1016/j.fishres.2021.106075

Liu, H., Ju, Y., Tamate, H., Wang, T., and Xing, X. (2021a). Phylogeography of Sika Deer (*Cervus Nippon*) Inferred from Mitochondrial Cytochrome-B Gene and Microsatellite DNA. *Gene* 772, 145375. doi:10.1016/j.gene.2020.145375

Liu, H., Liu, J., Yang, M., He, Y., and Wang, Y. (2020). SSR and SNP Polymorphic Feature Analysis Based on *Cheilinus Undulatus* Transcriptome. *Genomics Appl. Biol.* 39, 2451–2461. doi:10.13417/j.gab.039.002451

Liu, H., Zhang, Y., Wang, Z., Su, Y., and Wang, T. (2021c). Development and Application of EST-SSR Markers in *Cephalotaxus Oliveri* from Transcriptome Sequences. *Front. Genet.* 12, 759557. doi:10.3389/fgene.2021.759557

Liu, L.-X., Liu, Y.-G., and Xing, S.-C. (2014). An Analysis of Genetic Variability in Wild and Hatchery Populations of Swimming Crab (*Portunus Trituberculatus*) Using AFLP Markers. *Fish. Aquac. J.* 05, 1000104. doi:10.4172/2150-3508.1000104

Liu, L., Fu, Y., Zhu, F., Mu, C., Li, R., Song, W., et al. (2018a). Transcriptomic Analysis of *Portunus Trituberculatus* Reveals A Critical Role for WNT4 and WNT Signalling in Limb Regeneration. *Gene* 658, 113–122. doi:10.1016/j.gene.2018.03.015

Liu, M., Pan, J., Dong, Z., Cheng, Y., Gong, J., and Wu, X. (2019). Comparative Transcriptome Reveals the Potential Modulation Mechanisms of Estradiol Affecting Ovarian Development of Female *Portunus Trituberculatus*. *PLoS One* 14, e0226698. doi:10.1371/journal.pone.0226698

Liu, Q., Cui, F., Hu, P., Yi, G., Ge, Y., Liu, W., et al. (2018b). Using of Microsatellite DNA Profiling to Identify Hatchery-Reared Seed and Assess Potential Genetic Risks Associated with Large-Scale Release of Swimming Crab Portunus Trituberculatus in Panjin, Chinafiling to Identify Hatchery-Reared Seed and Assess Potential Genetic Risks Associated with Large-Scale Release of Swimming Crab *Portunus Trituberculatus* in Panjin, China. *Fish. Res.* 207, 187–196. doi:10.1016/j.fishres.2018.05.003

Liu, S., Sun, J., and Hurtado, L. A. (2013). Genetic Differentiation of Portunus Trituberculatus, the World's Largest Crab Fishery, Among its Three Main Fishing Areas. *Fish. Res.* 148, 38–46. doi:10.1016/j.fishres.2013.08.003

Liu, X., Xie, X., Liu, H., Nie, H., Ma, H., Li, D., et al. (2022). Population Genomic Evidence for Genetic Divergence in the Northwest Pacific Ark Shell (*Scapharca Broughtonii*). *Aquac. Rep.* 24, 101100. doi:10.1016/j.aqrep.2022.101100

Liu, Y.-G., Guo, Y.-H., Hao, J., and Liu, L.-X. (2012). Genetic Diversity of Swimming Crab (*Portunus Trituberculatus*) Populations from Shandong Peninsula as Assessed by Microsatellite Markers. *Biochem. Syst. Ecol.* 41, 91–97. doi:10.1016/j.bse.2011.12.024

Liu, Y., Liu, R., Ye, L., Liang, J., Xuan, F., and Xu, Q. (2009). Genetic Differentiation between Populations of Swimming Crab *Portunus Trituberculatus* along the Coastal Waters of the East China Sea. *Hydrobiologia* 618, 125–137. doi:10.1007/s10750-008-9570-2

Lu, C., Sun, Z., Xu, P., Na, R., Lv, W., Cao, D., et al. (2022). Novel Microsatellites Reveal Wild Populations Genetic Variance in Pike-Perch (*Sander lucioperca*) in China. *Aquac. Rep.* 23, 101031. doi:10.1016/j.aqrep.2022.101031

Lv, J., Gao, B., Liu, P., Li, J., and Meng, X. (2017). Linkage Mapping Aided by De Novo Genome and Transcriptome Assembly in Portunus Trituberculatus: Applications in Growth-Related QTL and Gene Identification. *Sci. Rep.* 7, 7874. doi:10.1038/s41598-017-08256-8

Lv, J., Li, R., Su, Z., Gao, B., Ti, X., Yan, D., et al. (2021). A Chromosome-Level Genome of *Portunus Trituberculatus* Provides Insights into its Evolution, Salinity Adaptation and Sex Determination. *Mol. Ecol. Resour.* 22, 1606–1625. doi:10.1111/1755–0998.13564

Lv, J., Liu, P., Gao, B., Wang, Y., Wang, Z., Chen, P., et al. (2014). Transcriptome Analysis of the *Portunus Trituberculatus: De Novo* Assembly, Growth-Related Gene Identification and Marker Discovery. *PLoS One* 9, e94055. doi:10.1371/journal.pone.0094055

Lv, J., Wang, Y., Gao, B., Li, J., and Liu, P. (2013). Identification of Type I Microsatellite Markers and Their Polymorphism in Portunus Trituberculatus. *J. Fish. China* 37, 816–822. doi:10.3724/SP.J.1231.2013.38343

Ma, H., Jiang, W., Liu, P., Feng, N., Ma, Q., Ma, C., et al. (2014). Identification of Transcriptome-Derived Microsatellite Markers and Their Association with the Growth Performance of the Mud Crab (*Scylla Paramamosain*). *PLoS One* 9, e89134. doi:10.1371/journal.pone.0089134

Ma, H., Li, L., Xiao, S., Zhang, Y., and Yu, Z. (2021). Microsatellite-Based Study of Population Genetics of *Crassostrea Hongkongensis* in Southern China. *Aquac. Rep.* 19, 100591. doi:10.1016/j.aqrep.2021.100591

Ma, K., Qiu, G., Feng, J., and Li, J. (2012). Transcriptome Analysis of the Oriental River Prawn, *Macrobrachium Nipponense* Using 454 Pyrosequencing for Discovery of Genes and Markers. *PLoS One* 7, e39727. doi:10.1371/journal.pone.0039727

Meng, W., Jiang, Y., Zhang, L., Yang, T., and Zhou, J. (2019). SSR Loci Information Analysis in the Transcriptome of Burbot (*Lota Lota*) Based on RNA-Seq. *Freshw. Fish.* 49, 10–14. doi:10.13721/j.cnki.dsyy.2019.06.002

Nagy, S., Poczai, P., Cernák, I., Gorji, A. M., Hegedűs, G., and Taller, J. (2012). PICcalc: An Online Program to Calculate Polymorphic Information Content

for Molecular Genetic Studies. *Biochem. Genet.* 50, 670–672. doi:10.1007/s10528-012-9509-1

Nguyen, C., Nguyen, T. G., Nguyen, L. V., Pham, H. Q., Nguyen, T. H., Pham, H. T., et al. (2016). De Novo assembly and Transcriptome Characterization of Major Growth-Related Genes in Various Tissues of *Penaeus monodon*. *Aquaculture* 464, 545–553. doi:10.1016/j.aquaculture.2016.08.003

Ni, S., Yang, Y., Liu, S., and Zhuang, Z. (2018). Microsatellite Analysis of *Patinopecten Yessoensis* Using Next-Generation Sequencing Method. *Prog. Fish. Sci.* 39, 107–113. doi:10.11758/yykxjz.20161209001

Pavan Kumar, P., Janakiram, T., and Bhat, K. V. (2020). Microsatellite Based DNA Fingerprinting and Assessment of Genetic Diversity in Bougainvillea Cultivars. *Gene* 753, 144794. doi:10.1016/j.gene.2020.144794

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research-Aan Update. *Bioinformatics* 28, 2537–2539. doi:10.1093/bioinformatics/bts460

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155, 945–959. doi:10.1093/genetics/155.2.945

Qin, Y., Shi, G., and Sun, Y. (2013). Evaluation of Genetic Diversity in *Pampus Argenteus* Using SSR Markers. *Genet. Mol. Res.* 12, 5833–5841. doi:10.4238/2013.November.22.10

Rousset, F. (2008). genepop'007: a Complete Re-implementation of the Genepop Software for Windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi:10.1111/j.1471-8286.2007.01931.x

Schlötterer, C., and Tautz, D. (1992). Slippage Synthesis of Simple Sequence DNA. *Nucl. Acids Res.* 20, 211–215. doi:10.1093/nar/20.2.211

Shan, B., Hui, M., Zhang, X., Liu, S., Cai, S., Song, N., et al. (2017). Genetic Effects of Released Swimming Crab (*Portunus Trituberculatus*) on Wild Populations Inferred from Mitochondrial Control Region Sequences. *Mitochondrial DNA Part A* 29, 856–861. doi:10.1080/24701394.2017.1373108

Singh, R. B., Mahenderakar, M. D., Jugran, A. K., Singh, R. K., and Srivastava, R. K. (2020). Assessing Genetic Diversity and Population Structure of Sugarcane Cultivars, Progenitor Species and Genera Using Microsatellite (SSR) Markers. *Gene* 753, 144800. doi:10.1016/j.gene.2020.144800

Song, L., Liu, P., Li, J., and Liu, Z. (2008). Analysis of Microsatellite Sequences in Genome of Crab, Portunus Trituberculatus. *J. Fish. Sci. China* 15, 738–744. doi:10.3321/j.issn:1005-8737.2008.05.004

Sun, H., Sun, C., Dong, J., Tian, Y., Hu, J., and Ye, X. (2019). Transcriptome Sequencing and Development and Application of Novel SSR Markers for *Siniperca chuatsi*. *Genomics Appl. Biol.* 38, 4413–4421. doi:10.13417/j.gab.038.004413

Tang, B., Zhang, D., Li, H., Jiang, S., Zhang, H., Xuan, F., et al. (2020). Chromosome-Level Genome Assembly Reveals the Unique Genome Evolution of the Swimming Crab (*Portunus Trituberculatus*). *GigaScience* 9, 1–10. doi:10.1093/gigascience/giz161

Tian, Z., Chen, A., Wu, Y., Chen, S., Zhang, Y., Cao, Y., et al. (2021). Bioinformatics Analysis of Microsatellite Sites in the RNA-Sequencing of <italic>Meretrix Meretrix</italic>. *Mf* 43, 160–167. doi:10.13233/j.cnki.mar.fish.20210422.001

Tulsani, N. J., Hamid, R., Jacob, F., Umretiya, N. G., Nandha, A. K., Tomar, R. S., et al. (2020). Transcriptome Landscaping for Gene Mining and SSR Marker Development in Coriander (*Coriandrum Sativum* L.). *Genomics* 112, 1545–1553. doi:10.1016/j.ygeno.2019.09.004

Vidya, V., Prasath, D., Snigdha, M., Gobu, R., Sona, C., and Maiti, C. S. (2021). Development of EST-SSR Markers Based on Transcriptome and its Validation in Ginger (*Zingiber Officinale* Rosc.). *PLoS ONE* 16, e0259146. doi:10.1371/journal.pone.0259146

Waiho, K., Ikhwanuddin, M., Afiqah-Aleng, N., Shu-Chien, A. C., Wang, Y., Ma, H., et al. (2022). Transcriptomics in Advancing Portunid Aquaculture: A Systematic Review. *Rev. Aquac.* 14, 1–25. doi:10.1111/raq.12689

Wang, H., Cui, Z., Wu, D., Guo, E., Liu, Y., Wang, C., et al. (2012). Application of Microsatellite DNA Parentage Markers in the Swimming Crab Portunus Trituberculatus. *Aquacult Int.* 20, 649–656. doi:10.1007/s10499-011-9493-1

Wang, Q., Liu, Y., Yan, L., Chen, L., and Li, B. (2021). Genome-Wide SNP Discovery and Population Genetic Analysis of *Mesocentrotus Nudus* in China Seas. *Front. Genet.* 12, 717764. doi:10.3389/fgene.2021.717764

Wang, S.-H., Zhang, C., Shang, M., Wu, X.-G., and Cheng, Y.-X. (2019b). Genetic Diversity and Population Structure of Native Mitten Crab (*Eriocheir Sensu Stricto*) by Microsatellite Markers and Mitochondrial COI Gene Sequence. *Gene* 693, 101–113. doi:10.1016/j.gene.2018.12.083

Wang, Z., Li, J., Hao, R., Adzigbli, L., and Deng, Y. (2019a). Characterization and Development of SSR Markers of *Pinctada Maxima* by RNA-Seq Approach. *Aquac. Rep.* 15, 100230. doi:10.1016/j.aqrep.2019.100230

Wang, Z., Li, X., Jiang, Y., Zuo, B., Si, Y., and Zheng, Y. (2018a). Population Genetic Analysis on *Eriocheir Sinensis* with the Red Shell Using Microsatellite Markers. *Chin. Fish. Qual. Stand.* 8, 34–41. doi:10.3969/j.issn.2095-1833.2018.03.005

Wang, Z., Sun, L., Guan, W., Zhou, C., Tang, B., Cheng, Y., et al. (2018b). *De Novo* Transcriptome Sequencing and Analysis of Male and Female Swimming Crab (*Portunus Trituberculatus*) Reproductive Systems During Mating Embrace (Stage Ii). *BMC Genet.* 19, 3. doi:10.1186/s12863-017-0592-5

Wong, L. L., Razali, S. A., Deris, Z. M., Danish-Daniel, M., Tan, M. P., Nor, S. A. M., et al. (2022). Application of Second-Generation Sequencing (SGS) and Third Generation Sequencing (TGS) in Aquaculture Breeding Program. *Aquaculture* 548, 737633. doi:10.1016/j.aquaculture.2021.737633

Xia, W., Xiao, Y., Liu, Z., Luo, Y., Mason, A. S., Fan, H., et al. (2014). Development of Gene-Based Simple Sequence Repeat Markers for Association Analysis in *Cocos nucifera*. *Mol. Breed.* 34, 525–535. doi:10.1007/s11032-014-0055-x

Xu, Q., and Liu, R. (2011). Development and Characterization of Microsatellite Markers for Genetic Analysis of the Swimming Crab, Portunus Trituberculatus. *Biochem. Genet.* 49, 202–212. doi:10.1007/s10528-010-9399-z

Xu, R., Wang, Z., Su, Y., and Wang, T. (2020). Characterization and Development of Microsatellite Markers in *Pseudotaxus Chienii* (Taxaceae) Based on Transcriptome Sequencing. *Front. Genet.* 11, 574304. doi:10.3389/fgene.2020.574304

Yang, W., Zheng, J., Jia, B., Wei, H., Wang, G., and Yang, F. (2018). Isolation of Novel Microsatellite Markers and Their Application for Genetic Diversity and Parentage Analyses in Sika Deer. *Gene* 643, 68–73. doi:10.1016/j.gene.2017.12.007

Yeh, F. C., Yang, R., and Boyle, T. (1999). *POPGENE Microsoft Window-Base Software for Population Genetic Analysis Version1.32*. Alberta, Canada: University of Alberta Center for International Forestry Research. A Quick User's Guide.

Yu, L., Zhu, X., Liang, J., Fan, J., and Chen, C. (2019). Analysis of Genetic Structure of Wild and Cultured Giant Freshwater Prawn (*Macrobrachium Rosenbergii*) Using Newly Developed Microsatellite. *Front. Mar. Sci.* 6, 323. doi:10.3389/fmars.2019.00323

Yue, H., Zhai, Q., Song, M., Ye, H., Yang, X., and Li, C. (2016). Development of Microsatellite Markers in *Cyprinus carpio* Var. *Singuonensis* Using Next-Generation Sequencing. *Freshw. Fish.* 46, 24–28. doi:10.13721/j.cnki.dsyy.2016.01.004

Yue, L., Wang, Y., Xian, W., and Zhang, H. (2022). Genetic Diversity and Population Structure of Portunustrituberculatus in Released and Wild Populations Based on Microsatellite DNA Markers from the Yangtze Estuary. *Diversity* 14, 374. doi:10.3390/d14050374

Zhai, Y., Wu, R., Niu, S., Shen, R., and Zhang, H. (2020). Development of Dinucleotide Microsatellite Loci for *Lateolabrax Maculates* Using Next Generation Sequencing Technology. *Genomics Appl. Biol.* 39, 507–513. doi:10.13417/j.gab.039.000507

Zhan, A., Hu, J., Hu, X., Lu, W., Wang, M., Peng, W., et al. (2008). Fast Identification of Scallop Adductor Muscles Using Species-specific Microsatellite Markers. *Eur. Food Res. Technol.* 227, 353–359. doi:10.1007/s00217-007-0728-3

Zhang, W., Lv, J., Lan, W., Gao, B., and Liu, P. (2022). Discovery of Sex-Determining Genes in *Portunus Trituberculatus*: A Comparison of Male and Female Transcriptomes During Early Developmental Stages. *Front. Mar. Sci.* 8, 811052. doi:10.3389/fmars.2021.811052

Zhao, Y., Zhu, X., Li, Z., Xu, W., Dong, J., Wei, H., et al. (2019). Genetic Diversity and Structure of Chinese Grass Shrimp, *Palaemonetes Sinensis*, Inferred from Transcriptome-Derived Microsatellite Markers. *BMC Genet.* 20, 75. doi:10.1186/s12863-019-0779-z

Zheng, X., Kuang, Y., Lü, W., Cao, D., and Sun, X. (2014). Transcriptome-derived EST-SSR Markers and Their Correlations with Growth Traits in Crucian Carp *Carassius auratus*. *Fish. Sci.* 80, 977–984. doi:10.1007/s12562-014-0782-2

Zhou, Q.-C., Shi, B., Jiao, L.-F., Jin, M., Sun, P., Ding, L.-Y., et al. (2019a). Hepatopancreas and Ovarian Transcriptome Response to Different Dietary

Soybean Lecithin Levels in *Portunus Trituberculatus*. *Comp. Biochem. Physiology Part D Genomics Proteomics* 31, 100600. doi:10.1016/j.cbd.2019.100600

Zhou, Q., Mu, K., Ni, Z., Liu, X., Li, Y., and Xu, L.-a. (2020b). Analysis of Genetic Diversity of Ancient *Ginkgo* Populations Using SSR Markers. *Industrial Crops Prod.* 145, 111942. doi:10.1016/j.indcrop.2019.111942

Zhu, J., Zhang, J., Jiang, M., Wang, W., Jiang, J., Li, Y., et al. (2021). Development of Genome-Wide SSR Markers in *Rapeseed* by Next Generation Sequencing. *Gene* 798, 145798. doi:10.1016/j.gene.2021.145798

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

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 22 August 2022
DOI 10.3389/fgene.2022.961009



OPEN ACCESS

EDITED BY
Yuzine Esa,
Putra Malaysia University, Malaysia

REVIEWED BY
Sheng Luan,
Yellow Sea Fisheries Research Institute (CAFS), China
Baohua Chen,
Xiamen University, China

*CORRESPONDENCE
Jianyong Liu,
liujy70@126.com

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

RECEIVED 03 June 2022
ACCEPTED 18 July 2022
PUBLISHED 22 August 2022

CITATION
Fu S and Liu J (2022), Genome-wide association study identified genes associated with ammonia nitrogen tolerance in Litopenaeus vannamei. *Front. Genet.* 13:961009. doi: 10.3389/fgene.2022.961009



# Genome-wide association study identified genes associated with ammonia nitrogen tolerance in *Litopenaeus vannamei*

Shuo Fu[1,2] and Jianyong Liu[1,2]*

[1]College of Fisheries, Guangdong Ocean University, Zhanjiang, China, [2]Guangdong Provincial Shrimp Breeding and Culture Laboratory, Guangdong Ocean University, Zhanjiang, China

Ammonia nitrogen tolerance is an economically important trait of the farmed penaeid shrimp *Litopenaeus vannamei*. To identify the genes associated with ammonia nitrogen tolerance, we performed an extreme phenotype genome-wide association study method (XP-GWAS) on a population of 200 individuals. The single nucleotide polymorphism (SNP) genotyping array method was used to construct the libraries and 36,048 SNPs were genotyped. Using the MLM, FarmCPU and Blink models, six different SNPs, located on SEQ3, SEQ4, SEQ5, SEQ7 and SEQ8, were determined to be significantly associated with ammonia nitrogen tolerance. By integrating the results of the GWAS and the biological functions of the genes, seven candidate genes (PDI, OZF, UPF2, VPS16, TMEM19, MYCBP2, and HOX7) were found to be associated with ammonia nitrogen tolerance in *L. vannamei*. These genes are involved in cell transcription, cell division, metabolism, and immunity, providing the basis for further study of the genetic mechanisms of ammonia nitrogen tolerance in *L. vannamei*. Further candidate gene association analysis in the offspring population revealed that the SNPs in the genes zinc finger protein OZF-like (OZF) and homeobox protein Hox-B7-like (HOX7) were significantly associated with ammonia nitrogen tolerance trait of *L. vannamei*. Our results provide fundamental genetic information that will be useful for further investigation of the molecular mechanisms of ammonia nitrogen tolerance. These associated SNPs may also be promising candidates for improving ammonia nitrogen tolerance in *L. vannamei*.

KEYWORDS

***Litopenaeus vannamei*, XP-GWAS, candidate genes, ammonia nitrogen tolerance, breeding**

## Introduction

*Litopenaeus vannamei* is naturally distributed along the Pacific coast of Central and South America. It is an important farmed penaeid shrimp that provides approximately 80% of the world's total penaeid shrimp output (FAO 2020). Semi-intensive and intensive cultivation methods are often adopted in the cultivation of *L. vannamei* to realize large-scale cultivation. With these methods, a large amount of feed is placed in

the aquaculture water, and the accumulation of residual bait and feces leads to a deterioration in water quality and an increase in the concentration of ammonia nitrogen (Romano and Zeng 2013).

Ammonia nitrogen in aquaculture water has two forms: non-ionic ammonia ($NH_3$) and ionic ammonia ($NH^{4+}$) (Emerson et al., 1975). Non-ionic ammonia can diffuse through the cell membrane and accumulate in the organs of aquatic animals (Kır et al., 2004; Cobo et al., 2014), causing organ damage and destroying the oxidant/antioxidant balance of aquatic animals, resulting in oxidative stress and an increased frequency of shrimp molting. Ultimately, this results in a loss of membrane integrity, reducing the immune capacity of shrimp and leading to death (Cheng et al., 2015; Jin J. et al., 2017; Li N. et al., 2018; Liang et al., 2019; Zhang et al., 2020).

To ensure sustainable development of *L. vannamei* aquaculture, it is important to improve the germplasm of *L. vannamei* and cultivate varieties with strong tolerance to high ammonia nitrogen. At present, selection of *L. vannamei* is generally based on population and family selection (Gjedrem 1985; Gjedrem and Baranski 2009; Wang 2013; Kong et al., 2020). De Donato et al. (2005) used a population breeding method to improve the average growth rate of *L. vannamei* by 35.5% after multiple generations of selection. Zhang et al. (2016) reported that the heritability of low dissolved oxygen tolerance of *L. vannamei* was also low at 0.07 ± 0.03. In terms of ammonia nitrogen resistance, Yuan et al. (2018) estimated the heritability of high ammonia nitrogen tolerance traits of *L. vannamei* at 7 and 14 weeks using the restricted maximum likelihood (REML) method, the traits were found to have low heritability (0.13 and 0.17). Li et al. (2016) also estimated the genetic parameters of ammonia nitrogen tolerance traits of *L. vannamei* larvae and found that the heritability of ammonia nitrogen tolerance traits of larvae was low.

Traditional breeding methods have long cycles and limited genetic progress (Montaldo and Castillo-Juárez 2017), especially for traits that cannot be measured directly and have low heritability. Marker-assisted breeding (MAS), which enables direct selection breeding of individuals with the aid of molecular markers tightly linked to the trait, is widely used in the genetic breeding of marine animals because it is associated with high genetic stability and discrimination (Ribaut and Hoisington, 1998; Lu et al., 2019). Yu, (2014) established a method to identify single nucleotide polymorphisms (SNPs) in *L. vannamei*, using high-throughput next-generation sequencing transcriptome data, 96,040 SNP markers of *L. vannamei* were successfully identified. Several studies have identified transcriptomic changes and differentially expressed genes in *L. vannamei* after high ammonia nitrogen stress. Xiao et al. (2019) identified several pathways and genes involved in ammonia nitrogen tolerance in *L. vannamei* based on comparative transcriptomic and metabolomic analyses of ammonia-tolerant and ammonia-sensitive *L. vannamei* families. Lu et al. (2016) identified 12 SNPs associated with ammonia nitrogen tolerance in *L. vannamei* using marker-trait correlation analyses.

Previous transcriptomic and metabolomic analyses have focused on the genetic bases of ammonia tolerance in shrimp. Quantitative trait loci (QTL) linkage mapping is an essential method for identifying relevant genes. Zeng et al. (2020) constructed a high-density genetic map of *L. vannamei* and mapped a QTL associated with ammonia nitrogen tolerance. Due to limitations of marker density, QTL analysis only identified one gene, LOC113809108, annotated as the ATP synthase g subunit.

Genome-wide association study (GWAS) can be used to identify functional genes in a genome. Until now, genes related to economic traits of important aquatic animals have been identified by GWAS, including catfish (*Ictalurus punctatus*) (Jin Y. et al., 2017; Li Y. et al., 2018), carp (*Cyprinus carpio*) (Zhou et al., 2018), and large yellow croaker (*Larimichthys crocea*) (Dong et al., 2019; Liu et al., 2020). Wang (2017) conducted a GWAS on growth and disease resistance traits of *L. vannamei*, identifying 52 SNPs significantly associated with body length, 47 SNPs associated with body weight, and 108 SNPs associated with *Vibrio* resistance. Jones et al. (2020) performed a GWAS for *L. vannamei* sex and found a QTL located on LG42.44 that was significantly associated with sex, but no related genes were annotated. Sun, (2021) performed GWAS analysis of *L. vannamei* growth and resistance to white spot syndrome virus (WSSV), found multiple significant loci and speculated that these two traits were controlled by micro-efficient polygenes. However, no study has investigated SNPs associated with ammonia nitrogen tolerance in *L. vannamei* using a GWAS.

Traditional GWAS analysis genotype-phenotype associations using large number of genotyped individuals, making GWAS an expensive analytical approach. The high cost of GWAS analysis is primarily related to the number of samples analyzed and the price of genotyping. To reduce costs, scientists have developed a GWAS approach in which only extreme samples are sequenced; this is referred to as an extreme phenotype genome-wide association study (XP-GWAS) (Yang et al., 2015). XP-GWAS is effective in reducing genotyping efforts, enabling low-cost and highly effective SNP screening (Dong et al., 2016; Wan et al., 2018).

In this study, we assessed ammonia nitrogen tolerance traits in *L. vannamei* using an XP-GWAS. Six different analysis methods were used to identify loci significantly associated with ammonia tolerance traits in *L. vannamei* and successfully mapped to potential genes associated with ammonia nitrogen tolerance. Finally, we confirmed that the use of XP-GWAS is feasible in *L. vannamei*, and is a cost-saving approach to genotyping.

# Materials and methods

## Material collection

All animal experiments were carried out following the National Institute of Health's Guide for the Care and Use of Laboratory Animals. The animal protocols were approved by the Animal Ethics Committee of Guangdong Ocean University (Zhanjiang, China). Shrimp were reared by the shrimp breeding company Guoxing Aquaculture Science and Technology Co., Ltd. in Zhanjiang City, Guangdong Province, P.R. China. In March 2020, we established 38 different families, and each family was cultured separately. After 1 month, 10 families were randomly selected and we moved 30 individuals from each family into the same pool for common environment breeding, each family was fluorescently labeled by visible implant elastomer. After 2 months of culture, 20 shrimps from each family were randomly selected for the high ammonia nitrogen stress experiment.

## High ammonia nitrogen challenge test

A population of 200 individuals was used for the challenge experiment (weight, 5.24 ± 2.07 g; body length, 77.49 ± 10.85 mm). An ammonia nitrogen stress concentration of 93 mg/L was used based on the results of the semi-lethal high ammonia nitrogen concentration of 96 h in preliminary experiments. For stress experiments, 20 shrimps from each family were placed in a 3 × 3 m test pool. After 7 days of suspension, the concentration of ammonia nitrogen was adjusted to 93 mg/L by adding analytical grade ammonium chloride.

During the experiment, water was replaced daily, the water temperature was maintained at 26 ± 2°C, the pH was maintained at 8.1 ± 0.2, the salinity was maintained at 30 ± 1, and the dissolved oxygen was maintained at 6–8 mg/L. Deaths were observed each hour and the survival time was recorded. Shrimp were considered dead if they lay on the pool floor with no touch response. After the time of death was recorded, the muscle tissues of each shrimp were preserved individually in absolute ethanol for subsequent DNA extraction. The experiment was concluded until all shrimp had died.

## DNA extraction

To reduce costs and improve efficiency, we chose extreme individuals for experiments based on Dong et al. (2016). We classified individuals according to survival time and selected 30% of the individuals who died first (Sensitive group) and 30% of the individuals who died last (Tolerance group) for DNA extraction and subsequent analysis. DNA was extracted from the muscle of each shrimp ($n$ = 120) with the EasyPure® Genomic DNA Kit (TRANSGEN BIOTECH) in the Nanhai Economic Shrimp Breeding and Culture Laboratory. The quality and concentration of extracted DNA was determined by agarose testing and a NanoDrop 2000 UV spectrophotometer (Thermo Fisher). DNA was stored at −20°C until further use.

## Phenotype and genotype statistics

Resistance to high ammonia nitrogen was measured as the number of hours to death. We used the Genobaits® Prawn 40K Panel developed by MOLBREEDING® (Hebei, China) as the genotyping method: it integrates the relevant SNPs including WSSV resistance, ammonia nitrogen tolerance, and feed conversion efficiency, SNPs were filtered using plink v1.9 (Chang et al., 2015). SNPs were filtered according to the following parameters: "plink --file GWAS --geno 0.1 --maf 0.01 --hwe 1e-5 --recode --out SNP".

## Statistical methods

We used six different models to implement GWAS, including General Linear Model (GLM), Mixed Linear Model (MLM), Compressed MLM (CMLM), Settlement of MLM Under Progressively Exclusive Relationship (SUPER), Fixed and random model Circulating Probability Unification (FarmCPU) and Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (Blink). All the GWAS analyses used the GAPIT R package (Lipka et al., 2012). MLM can be described as: $y = X\alpha + Q\beta + K\mu + e$, where y is the vector of phenotypic records (survival time after stress), X is the genotype matrix, α is the genotype effect vector, Q is the fixed effect matrix, β is the fixed effect vector including population structure and body length, K is the random effect matrix (kinship matrix), μ is the random effect vector, and e is the residual vector (Zhang et al., 2010). GLM only considers the effect of the genotype matrix and does not include fixed and random effects. CMLM clusters individuals into groups and fits genetic values of groups as random effects in the model to avoid false negatives in MLM (Li et al., 2014). In SUPER, only the associated genetic markers are used as pseudo Quantitative Trait Nucleotides (QTNs) to derive kinship, which can improve the statistical power compared to using overall kinship from all markers (Wang et al., 2014). FarmCPU and Blink are multilocus models that use a modified MLM method, multiple loci linear mixed model (MLMM), incorporate multiple markers simultaneously as covariates in a stepwise MLM to partially remove the confounding between testing markers and kinship (Liu et al., 2016; Huang et al., 2019; ; Kaler et al., 2019). The –log10 ($p$-value) of each SNP across the genome was calculated to illustrate the GWAS results. The threshold $p$-value for genome-wide significance was calculated using Bonferroni correction based

TABLE 1 The primers designed for the identification of SNPs in candidate gene.

| Primer ID | Primer sequence (5′-3’) | Length (bp) |
|---|---|---|
| HOX7 | Forward: TAGCACACCGCATTTCCTCA | 895 |
| | Reverse: CTTTCCCTATCCAACCAGCA | |
| OZF | Forward: CCGATTGGATGCCTTTGGA | 995 |
| | Reverse: GCTTCTCTGTTGTATGTGCTCTTT | |

on the estimated number of independent markers. Considering the marker number and the number of genotyping populations, loci that ranked in the top 200 in the results of all GWAS model analyses were finally included as significant SNPs.

## Candidate gene annotation

We annotated candidate genes around the significant SNPs based on the reference genome of *L. vannamei* (NCBI assembly ID: ASM378908v1). To do this, we defined a region that was expected to contain causative genes. Because the SNPs detected in this study are on different scaffolds, we assumed they have independent effects on investigated traits, so we defined an independent region that 300 kb upstream and downstream of each SNP to search candidate genes. To annotate the genes, we used the BLAST + tool to generate a sequence alignment using the non-redundant protein database (NCBI), the reference genome used in this study is ASM378908v1.

## Candidate gene association analysis

To validate the association of SNPs with high ammonia nitrogen tolerance traits in *L.vannamei*, we performed High ammonia nitrogen stress on a progeny population with 150 shrimp. Based on the results of the semi-lethal high ammonia nitrogen concentration of 48 h in preliminary experiments, ammonia nitrogen stress concentration was set to 383 mg/L. SNPs of gene HOX and OZF were examined by direct sequencing of PCR products (Sangon Biotech, Shanghai), information of primers designed for this procedure is listed in Table 1, all primers were synthesized by Shanghai Sangon Biotech. Comparison of the means of survival time among different genotypes was conducted by $\chi^2$-test using SPSS version 24.0. Significance level for the analysis was specified at $p < 0.05$.

# Results

## Descriptive statistics of phenotypic values

The average body weight and body length of individual shrimp were 5.24 ± 2.07 g and 77.49 ± 10.85 mm, respectively. Mortality was observed beginning at 14.5 h after the start of stress and lasted until 136 h (Figure 1), the median lethal time ($LT_{50}$) was 98 h. Statistical analysis revealed that the mortality data of *L. vannamei* were approximately normally distributed. Besides, the range of survival time in the sensitive and tolerant groups was 14.5–88 h and 110–136 h, respectively.

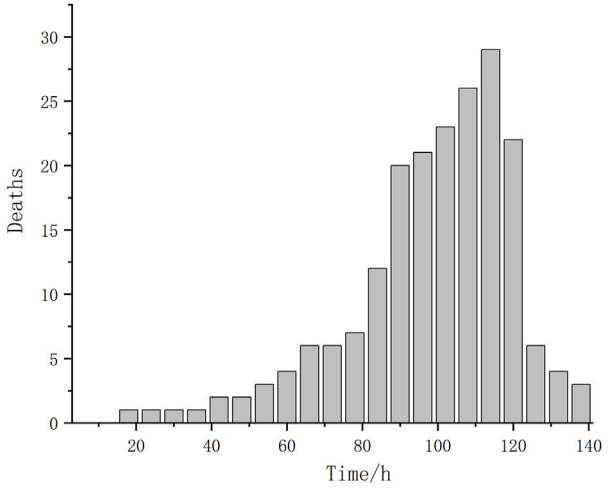



FIGURE 1
The death numbers of shrimp in different time after ammonia nitrogen stress.

## Genotyping and genome-wide association study analysis

We assessed the SNP quality and excluded 2,845 SNPs with a call rate that lower than 90%, 5,569 SNPs with minimum allele frequency that lower than 0.01, and 945 SNPs with significant deviation from Hardy–Weinberg equilibrium ($p < 0.00001$). After filtering, 36,048 quality compliant SNP loci were retained for GWAS analysis. We evaluated these eight models for false positives and false negatives based on the Q-Q plots (Figure 2). The results show that model MLM, CMLM, FarmCPU and Blink can control false positives and false negatives, whereas GLM and SUPER exhibit severe false

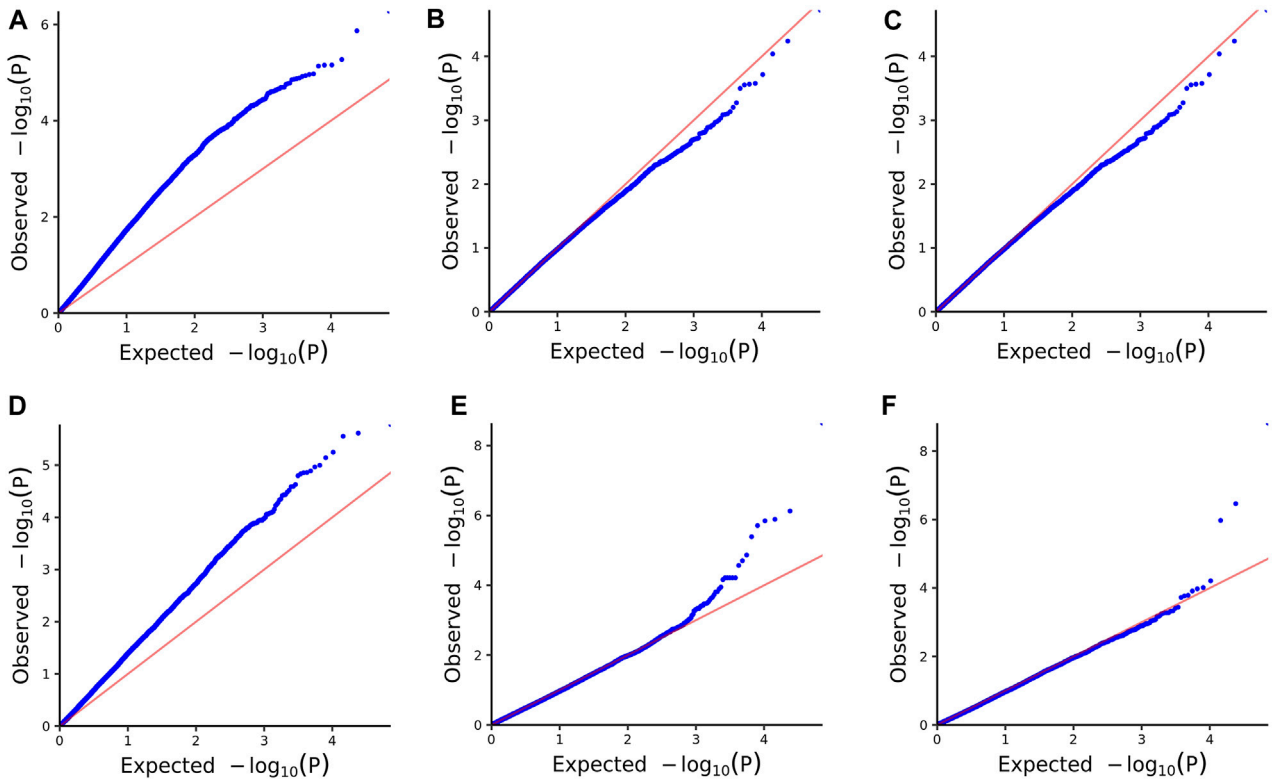



FIGURE 2
Quantile-quantile (QQ) plots of the six models. **(A)** GLM. **(B)** MLM. **(C)** CMLM. **(D)** SUPER. **(E)** FarmCPU. **(F)** Blink.

TABLE 2 Markers associated with ammonia nitrogen tolerance of *L. vannamei*.

| Number | SNP position | SNP | *p*-value (MLM) | *p*-value (Blink) | *p*-value (FarmCPU) |
|---|---|---|---|---|---|
| 1 | SEQ7 | A/C | 6.95E-06 | 1.29E-08 | 1.25E-10 |
| 2 | SEQ8 | A/C | 5.47E-05 | 5.45E-08 | 9.91E-09 |
| 3 | SEQ5 | T/C | 1.90E-04 | 8.93E-04 | 0.0019 |
| 4 | SEQ4 | A/G | 3.17E-04 | 0.0017 | 9.15E-04 |
| 5 | SEQ3 | A/G | 0.0011 | 0.0032 | 0.0013 |
| 6 | SEQ5 | A/C | 0.0045 | 4.24E-04 | 0.0066 |

positives. Considering that the results of CMLM are completely consistent with MLM, we only Integrated the results of MLM, FarmCPU and Blink, thus a total of six SNPs were selected (Table 2). The GWAS results are presented as Manhattan plots (Figure 3), and a Venn diagram was plotted to represent the intersection of the top 200 significant loci from these three models (Figure 4). The details of these top 200 significant loci can be found in Supplementary Data Sheet S1. Because the currently available *L. vannamei* genome is still at the scaffold level, with a large number of scaffolds, we concatenated the scaffolds into nine sequences (SEQ1–9). The six associated SNPs are scattered over SEQ5, SEQ4, SEQ3, SEQ7, and SEQ8.

## Candidate genes

The significant SNPs identified by the GWAS were used as probes to find the closest candidate gene up- and downstream according to its position on the genome. Seven candidate genes were found in the vicinity of the six SNPs detected

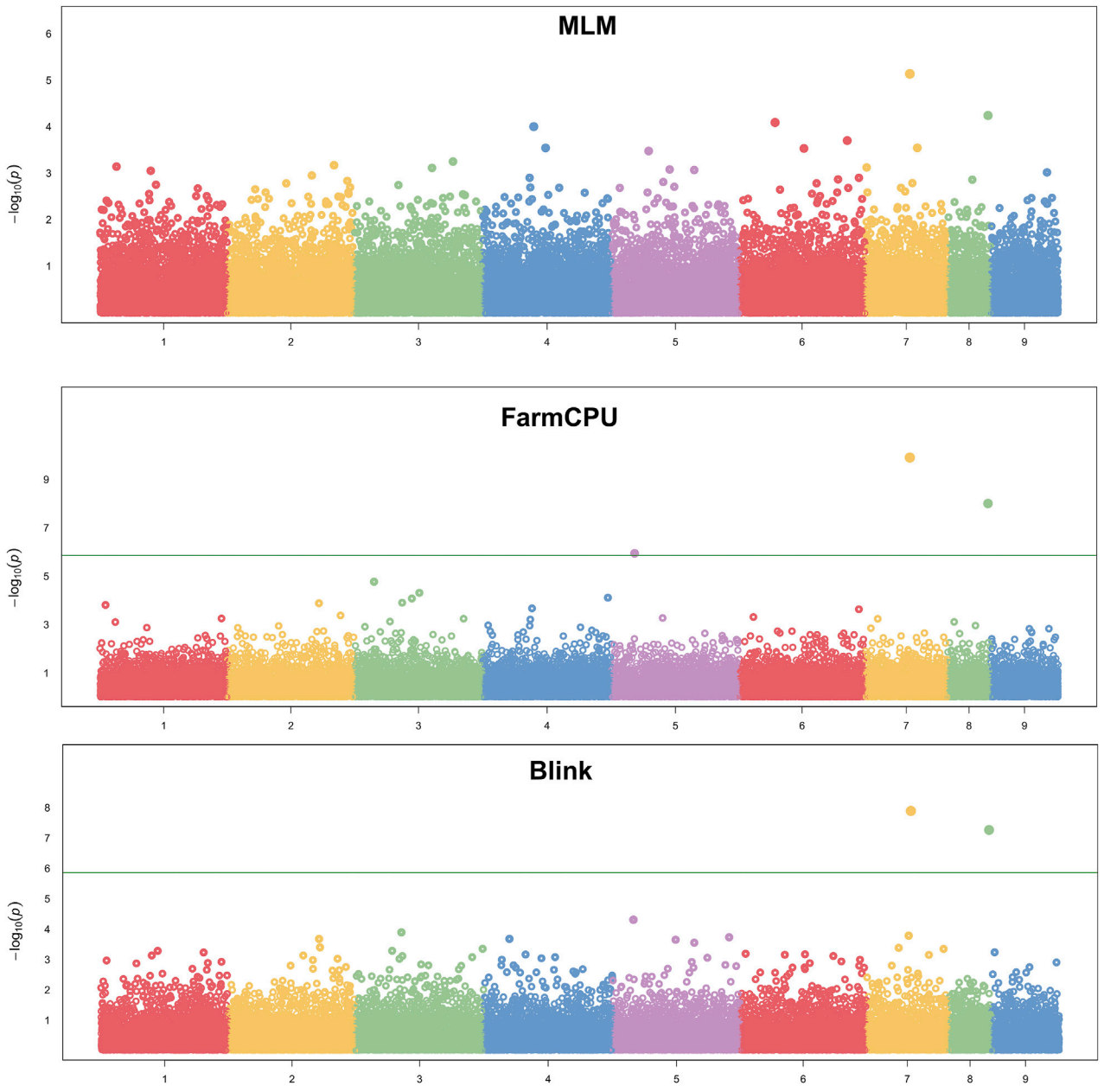



FIGURE 3
Manhattan plot of $-\log_{10} P$ for the GWAS of ammonia nitrogen tolerance.

(Table 3): PDI (protein disulfide-isomerase-like), OZF (zinc finger protein OZF-like), HOX7 (homeobox protein Hox-B7-like), RENT2 (regulator of nonsense transcripts 2-like), VPS16 (vacuolar protein sorting associated protein 16 homolog), TMEM19 (transmembrane protein 19-like) and MYCBP2 (E3 ubiquitin protein ligase mycbp 2-like). Analysis of these genes revealed that they are related to immune defense, apoptosis, growth, osmoregulation, and molting.

## Association analyses of candidate genes

Based on the *p*-value and the annotated gene function of the identified markers, we selected gene homeobox protein Hox-B7-like and zinc finger protein OZF-like for candidate genes association analyses in a progeny population. The average body weight and body length of individual shrimp in this progeny population were 15.73 ± 6.02 g and 117.70 ± 14.67 mm, respectively. Survival time ranged from 10.5–180 h

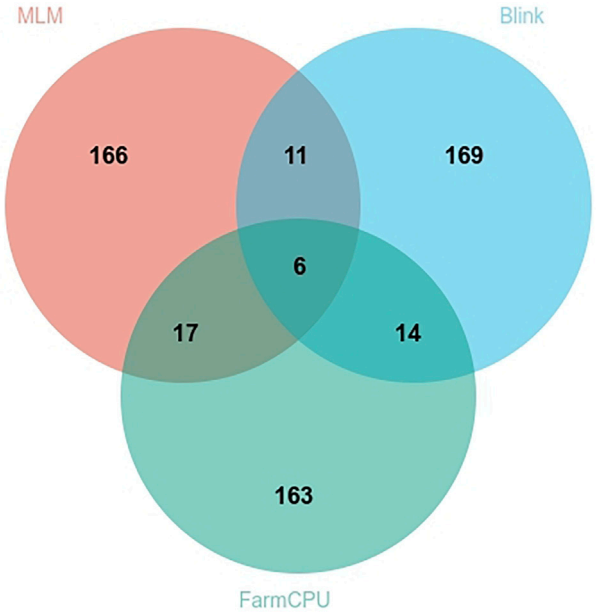



FIGURE 4
Venn plot based on the top 200 significant sites obtained from different statistical models.

and the median lethal time ($LT_{50}$) was 76.5 h. After genotyping these SNPs of two genes in the progeny population, SNPs of HOX7 and OZF were significantly associated with ammonia nitrogen tolerance (Table 4). In the HOX7 A>C locus, the distribution of the two different genotypes, AA and AC, showed significant difference in survival time after high ammonia nitrogen stress, with 76.5 h and 92.7 h, respectively, and the individuals with AA genotype had a significantly higher survival time than those with the AC genotype. In the OZF T>C locus, the mean survival times of individuals with genotypes TT, TC and CC were 77.6 h, 85.3 h and 55.7 h, respectively, individuals with the TT and TC genotypes survived significantly longer than those with the CC genotype.

## Discussion

Several studies have investigated genes associated with ammonia nitrogen resistance in *L. vannamei*, but many of these were based on transcriptome screening of SNPs for association and QTL linkage analysis (Xiao et al., 2019). Transcriptome analysis can be used to find a large number of differentially expressed genes, but it is difficult to discern which genes are responsible for the trait of interest. QTL analysis can only detect a few loci for economic traits, and thus lacks precision. In *L. vannamei*, association analysis with ammonia nitrogen resistance based on genetic variation within the genome has not yet been reported. To our knowledge, this is the first report about GWAS of ammonia nitrogen tolerance in *L. vannamei*. A number of SNPs significantly associated with ammonia nitrogen tolerance were identified by using MLM, FarmCPU and Blink models. However, the *p*-value of these SNPs is relatively low, which could confirm previous speculations that ammonia nitrogen tolerance is highly polygenic and may be controlled by many genes with moderate to low effects (Lu et al., 2016; Zeng et al., 2020).

TABLE 3 Candidate genes found by GWAS of ammonia nitrogen tolerance.

| Number | Name | Annotation | SNP position | Distance (kb) | Function |
|---|---|---|---|---|---|
| 1 | MYCBP2 | E3 ubiquitin-protein ligase MYCBP2-like | SEQ7 | U111.7 | developmental |
| 2 | PDI | protein disulfide-isomerase-like | SEQ8 | U20.8 | immune response |
| 3 | OZF | zinc finger protein OZF-like | SEQ5 | exon | DNA replication |
| 4 | UPF2 | regulator of nonsense transcripts 2-like | SEQ4 | U147.5 | immune response |
| 5 | VPS16 | vacuolar protein sorting-associated protein 16 homolog | SEQ3 | U293.9 | metabolism |
| 6 | TMEM19 | transmembrane protein 19-like | | D194.4 | cell transcription |
| 7 | HOX7 | homeobox protein Hox-B7-like | SEQ5 | intronic | cell division |

TABLE 4 Association analysis result of SNPs in the genes HOX7 and OZF.

| Genes | SNP | Genotype | ST (h) | Number | $\chi^2$ | *p*-value |
|---|---|---|---|---|---|---|
| HOX7 | CCCTTGTTCACTATAGACCTTCGTA [A/C]GGCTCACTGACACTATGTGGTAGAT | A/A | 76.5 | 122 | 6.940 | 0.008 |
| | | A/C | 92.7 | 28 | | |
| OZF | AAACGTTTTGCATGTGACGTGTGTG [T/C]CAAGAAATTCTCTTCTCGGAGTAAG | T/T | 77.6 | 64 | 9.414 | 0.009 |
| | | T/C | 85.3 | 75 | | |
| | | C/C | 55.7 | 11 | | |

Generally, Bonferroni correction is used to avoid false positives. However, the Bonferroni-corrected *p*-value is too strict in GWAS (Spencer et al., 2009; Hong and Park 2012). In the present study, the genome-wide significance threshold was approximately $1.38 \times 10^{-6}$ after Bonferroni correction, no SNPs exceeding the threshold were found in MLM, nor were only 2 SNPs exceeding the threshold found in FarmCPU and Blink. Considering the small sample size and large number of SNPs, the corrected significant threshold is too strict to identify candidate markers. Similar results were reported from a GWAS of growth and disease resistance traits in *L. vannamei* using reduced-representation genome sequencing (Wang 2017; Wang et al., 2019a; Wang et al., 2019b; Yu et al., 2019; Lyu et al., 2021). However, Sun (2021) performed a GWAS of *L. vannamei* growth and WSSV resistance by resequencing and found many SNPs that reached the threshold after Bonferroni correction. It is possible that because of the large number of repetitive sequences in the genome of *L. vannamei*, methods such as simplified genome sequencing and gene chips cannot allow accurate mapping due to low marker density, high-density markers covering the whole genome range are required for detection. Therefore, we widened the range of significant SNP loci for screening candidate genes to the top 200 significant. At the same time, in order to improve the rigor of the results, only the top 200 significant SNPs in all three analytical models (MLM, FarmCPU and Blink) simultaneously were considered to be significantly associated with high ammonia nitrogen tolerance in this study. MLM is one kind of single SNP analysis methods, some studies have found that multilocus models performed better than single locus models (Liu et al., 2016; Wen et al., 2018), the result of the Q-Q plot in this study also shows that the MLM model controls false positives while causing false negatives. However, as the most commonly used analysis model in current GWAS research, we still considered the results obtained by MLM analysis, and finally a total of 6 SNPs were found.

The SNPs related to ammonia nitrogen tolerance found in this study mainly have functions related to cell transcription, cell division, metabolism, and immunity. Ammonia nitrogen tolerance may stimulate physiological reactions in *L. vannamei*, triggering gene transcription, replication, and the immune response. Zinc finger protein OZF is a member of zinc finger family and function to bind DNA, RNA, protein, and lipid substrates (Andrea, 2001; Hall, 2005; Sonia et al., 2020). Zuo et al. (2018) found that another member of the zinc finger family, single C4-containing zinc finger protein can affect DNA replication and positively regulate the expression of various antimicrobial peptides, thus indirectly participating in the antibacterial response of *L. vannamei*. HOX7 is a member of the HOX gene family, which encodes transcription factors that regulate cell division (Zou and Jiang, 2008). HOX genes specify cell fates in animal embryos, and influence body weight gain (Lee et al., 2014). UPF2 is a conserved nonsense-mediated mRNA decay factor. Nonsense-mediated mRNA decay, also called mRNA surveillance, is an important pathway used by all organisms to degrade mRNAs that prematurely terminate translation, and consequently eliminate the production of aberrant proteins that could be harmful. In the UPF trimeric complex, UPF2 and UPF3b cooperatively stimulate both ATPase and RNA helicase activities of UPF1 (Johnson et al., 2019). VPS11, a component of the vacuolar protein sorting (VPS) subunit C, composed of VPS11, VPS18, VPS16, and VPS33A proteins, is involved in the tethering of endosomes, lysosomes, and autophagosomes (Bröcker et al., 2012; Li et al., 2012). The TMEM19 gene is a novel gene with no known function. Based on current reports of transmembrane (TMEM) protein family members, TMEM proteins are involved in intercellular and intracellular signal transduction and immune-related diseases, as well as many physiological processes, such as forming ion channels in the plasma membrane, activating signal transduction pathways, and mediating cell chemotaxis, adhesion, apoptosis, and autophagy (Li, 2007). However, no gene directly related to ammonia nitrogen tolerance was identified near the SNP sites screened in this study. This might be due to the sequencing method used. LOC113809108 (ATP synthase g subunit), a gene previously identified to be associated with ammonia nitrogen tolerance through QTL mapping studies, was not identified in this study, possibly due to the small marker density of the microarray. Therefore, further screening in large populations by re-sequencing will be necessary.

In this study, one SNP A>C in the intron of gene HOX7 and one SNP T>C in the exon of gene OZF were screened. Qian et al. (2013) found one SNP which was a synonymous mutation in the coding region of cathepsin CTSL gene of *L. vannamei*. After analysis, the SNP site had a significant effect on the growth characteristics of *L. vannamei*. Chen et al. (2016) analyzed single nucleotide polymorphisms of gene CAT and its correlation with low hemolytic oxygen tolerance traits in *L. vannamei*, found that one SNP, g.155 A>G, belonging to the synonymous mutation Gln→Gln, was screened in the CAT sequence and significantly associated with low hemolytic oxygen tolerance traits in *L. vannamei*. In our study, the SNP of HOX7 is a synonymous mutation Lys→Lys, the SNP of OZF is a nonsynonymous mutation Leu→Pro. As several studies have described, nonsynonymous SNPs change their encoded amino acid sequence, thus affecting gene functions and interactions (Wang and Moult, 2001; Ng and Henikoff, 2006; Teng et al., 2008). Synonymous SNPs did not change the encoded amino acid sequences, but base changes may indirectly affect the structure of the original gene, thereby affecting selective clipping and clipping efficiency of the gene, altering mRNA folding and the protein synthesis rate, ultimately affecting mRNA stability and the translational process of the gene (Liao and Lee, 2010). In addition, the

other four SNPs located in the intergenic region were not studied in this paper, because it is a very complicated process for intergenic SNPs to affect gene function. However, in the genome, the number of SNPs located in the intergenic region is much more than that located inside the genes, and some intergenic SNPs located in the promoter or enhancer even have more substantial effects (Hoogendoorn et al., 2004; Mishiro et al., 2009; Wagschal et al., 2015), so in the follow-up papers, we will continue to explore the functions of intergenic SNPs.

Compared with methods such as transcriptome analysis and QTL mapping, GWAS are costly because they require SNP genotyping across the genome of hundreds of individuals (Korte and Farlow 2013). The number of individuals to genotype can be reduced to reduce costs, although this also decreases accuracy. Some research teams have investigated extreme phenotypes in maize and large yellow croaker using XP-GWAS (Yang et al., 2015; Dong et al., 2016). GWAS of economic traits in the large yellow croaker with different numbers of extreme phenotypes revealed that 60% of the extreme phenotypic samples gave similar results as a GWAS with whole phenotypes, thus saving 40% of the genotyping and DNA extraction costs (Wan et al., 2018). We also performed XP-GWAS using 60% of extreme phenotypic samples and found several significant SNPs associated with ammonia nitrogen tolerance. Although the *p*-values of these SNPs did not reach the Bonferroni-corrected *p*-value threshold in MLM, the results were similar to those obtained by Wang et al. (2019a) in their GWAS of the resistance of *L. vannamei* against *Vibrio parahaemolyticus*, which involved analysis of whole individuals.

Quantitative traits are often influenced by multiple genes with small effects, so the benefit of conducting MAS depends on the effect of each SNP (Meuwissen et al., 2001). Previous GWAS have shown that resistance traits are not controlled by a major effect of one QTL, but by several polygenic genes with minor effects (Liu et al., 2015; Correa et al., 2017; Jin Y. et al., 2017). The results of this study also indicate that there may not be a major QTL that contributes to ammonia nitrogen tolerance of *L. vannamei*. Thus, the implementation of MAS may not be successful. Similar to our findings, Correa et al. (2015) performed a GWAS to assess the resistance of Atlantic salmon to *Piscirickettsia salmonis*; they concluded that it was due to a micro effect polygenic trait with low *p*-values for SNPs associated with this trait, suggesting that genomic selection will be a more efficient approach to such traits. Molecular information from genotyped SNPs may be incorporated in breeding programs through the application of genomic selection (Goddard and Hayes, 2007), where effects of all genotyped SNPs are included without the need to surpass a determined threshold of significance (Meuwissen et al., 2001). Such an approach should be evaluated to determine the usefulness of genotyped SNPs for ammonia nitrogen tolerance.

## Data availability statement

The data presented in the study are deposited in the NCBI repository, accession number: PRJNA850509.

## Author contributions

SF: Conceptualization, formal analysis, investigation, validation, writing—original draft, writing—review and editing. JL: Conceptualization, resources, supervision, project administration.

## Funding

This research was supported by the National key R&D plan "blue granary science and technology innovation" key special project in 2020 (2020YFD0900205); the Project of 2019 Annual Guangdong Provincial Special Financial Fund of Nanhai Economic Shrimp Breeding and Culture Laboratory (No. 2319412525).

## Conflict of interest

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.961009/full#supplementary-material

# References


Andrea, H., and HArtwig, A. (2001). Zinc finger proteins as potential targets for toxic metal ions: Differential effects on structure and function. *Antioxid. Redox Signal.* 3, 625–634. doi:10.1089/15230860152542970

Bröcker, C., Kuhlee, A., Gatsogiannis, C., Balderhaar, H. J. K., Hönscher, C., Engelbrecht-Vandré, S., et al. (2012). Molecular architecture of the multisubunit homotypic fusion and vacuole protein sorting (HOPS) tethering complex. *Proc. Natl. Acad. Sci. U. S. A.* 109, 1991–1996. doi:10.1073/pnas.1117797109

Chang, C., Chow, C., Tellier, L., Vattikuti, S., Purcell, S., and Lee, J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* 4, 7. doi:10.1186/s13742-015-0047-8

Chen, X., Liu, J., Zhang, J., Yuan, R., and Qian, J. (2016). Single nucleotide polymorphisms in catalase gene and their association with resistant hypoxia traits in *Litopenaeus vannamei*. *J. Zhanjiang Ocean. Univ.* 36 (06), 16–20. doi:10.3969/j.issn.1673-9159.2016.06.003

Cheng, C., Yang, F., Ling, R., Liao, S., Miao, Y., Ye, C., et al. (2015). Effects of ammonia exposure on apoptosis, oxidative stress and immune response in pufferfish (*Takifugu obscurus*). *Aquat. Toxicol.* 164, 61–71. doi:10.1016/j.aquatox.2015.04.004

Cobo, M. L., Sonnenholzner, S., Wille, M., and Sorgeloos, P. (2014). Ammonia tolerance of *Litopenaeus vannamei* (Boone) larvae. *Aquac. Res.* 45, 470–475. doi:10.1111/j.1365-2109.2012.03248.x

Correa, K., Lhorente, J. P., Bassini, L., López, M. E., Di Genova, A., Maass, A., 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, 61–65. doi:10.1016/j.aquaculture.2016.04.008

Correa, K., Lhorente, J. P., López, M. E., Bassini, L., Naswa, S., Deeb, N., et al. (2015). Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar L.*) chromosomes. *BMC Genomics* 16, 854. doi:10.1186/s12864-015-2038-7

De Donato, M., Manrique, R., Ramirez, R., Mayer, L., and Howell, C. (2005). Mass selection and inbreeding effects on a cultivated strain of Penaeus (*Litopenaeus vannamei*) in Venezuela. *Aquaculture* 247, 159–167. doi:10.1016/j.aquaculture.2005.02.005

Dong, L., Han, Z., Fang, M., Xiao, S., and Wang, Z. (2019). Genome-wide association study identifies loci for body shape in the large yellow croaker (*Larimichthys crocea*). *Aquac. Fish.* 4, 3–8. doi:10.1016/j.aaf.2018.05.001

Dong, L., Xiao, S., Chen, J., Wan, L., and Wang, Z. (2016). Genomic selection using extreme phenotypes and pre-selection of SNPs in large yellow croaker (*Larimichthys crocea*). *Mar. Biotechnol.* 18, 575–583. doi:10.1007/s10126-016-9718-4

Emerson, K., Russo, R. C., Lund, R. E., and Thurston, R. V. (1975). Aqueous ammonia equilibrium calculations: Effect of pH and temperature. *J. Fish. Res. Bd. Can.* 32, 2379–2383. doi:10.1139/f75-274

FAO (2020). *The state of world Fisheries aquaculture 2020*. Rome, Italy: Food and Agriculture Organization of the United Nations.

Gjedrem, T., and Baranski, M. (2009). *Selective breeding in aquaculture: An introduction*, 10. Springer Netherlands, 570–572.

Gjedrem, T. (1985). Improvement of productivity through breeding schemes. *GeoJournal* 10, 233–241. doi:10.1007/bf00462124

Goddard, M. E., and Hayes, B. J. (2007). Genomic selection. *J. Anim. Breed. Genet.* 6, 323–330. doi:10.1111/j.1439-0388.2007.00702.x

Hall, T. M. (2005). Multiple modes of RNA recognition by zinc finger proteins. *Curr. Opin. Struct. Biol.* 15, 367–373. doi:10.1016/j.sbi.2005.04.004

Hong, E. P., and Park, J. W. (2012). Sample size and statistical power calculation in genetic association studies. *Genomics Inf.* 10, 117–122. doi:10.5808/GI.2012.10.2.117

Hoogendoorn, B., Coleman, S. L., Guy, C. A., Smith, S. K., O'Donovan, M. C., and Buckland, P. R. (2004). Functional analysis of polymorphisms in the promoter regions of genes on 22q11. *Hum. Mutat.* 24, 35–42. doi:10.1002/humu.20061

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). Blink: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8. doi:10.1093/gigascience/giy154

Jin, J., Wang, Y., Wu, Z., Hergazy, A., Lan, J., Zhao, L., et al. (2017). Transcriptomic analysis of liver from grass carp (*Ctenopharyngodon idellus*) exposed to high environmental ammonia reveals the activation of antioxidant and apoptosis pathways. *Fish. Shellfish Immunol.* 63, 444–451. doi:10.1016/j.fsi.2017.02.037

Jin, Y., Zhou, T., Geng, X., Liu, S., Chen, A., Yao, J., et al. (2017). A genome-wide association study of heat stress-associated SNPs in catfish. *Anim. Genet.* 48, 233–236. doi:10.1111/age.12482

Johnson, J. L., Stoica, L., Liu, Y., Zhu, P. J., Bhattacharya, A., Buffington, S. A., et al. (2019). Inhibition of upf2-dependent nonsense-mediated decay leads to behavioral and neurophysiological abnormalities by activating the immune response. *Neuron* 104, 665–679. doi:10.1016/j.neuron.2019.08.027

Jones, D. B., Nguyen, H. T., Khatkar, M. S., Simma, D. B., Jerry, D. R., Raadsma, H. W., et al. (2020). The identification of a major sex QTL in the white-leg shrimp, *Litopenaeus vannamei*. *Aquaculture* 529, 735673. doi:10.1016/j.aquaculture.2020.735673

Kaler, A. S., Gillman, J. D., Beissinger, T., and Purcell, L. C. (2019). Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Front. Plant Sci.* 10, 1794. doi:10.3389/fpls.2019.01794

Kır, M., Kumlu, M., and Eroldoğan, O. T. (2004). Effects of temperature on acute toxicity of ammonia to Penaeus semisulcatus juveniles. *Aquaculture* 241, 479–489. doi:10.1016/j.aquaculture.2004.05.003

Kong, J., Luan, S., Tan, J., Sui, J., Luo, K., Li, X., et al. (2020). *Progress of study on penaeid shrimp selective breeding*, 50. Periodical of Ocean University of China, 81–97. doi:10.16441/j.cnki.hdxb.20200033

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* 9, 29. doi:10.1186/1746-4811-9-29

Lee, H. M., Rim, H. K., Seo, J. H., Kook, Y. B., Kim, S. K., Oh, C. H., et al. (2014). HOX-7 suppresses body weight gain and adipogenesis-related gene expression in high-fat-diet-induced obese mice. *BMC Complem. Altern. M.* 14, 505. doi:10.1186/1472-6882-14-505

Li, C., Weng, S., Chen, Y., Yu, X., Lu, L., Zhang, H., et al. (2012). Analysis of *Litopenaeus vannamei* transcriptome using the next-generation DNA sequencing technique. *PLoS One* 7, e47442. doi:10.1371/journal.pone.0047442

Li, M., Liu, X., Bradbury, P., Yu, J., Zhang, Y., Todhunter, R. J., et al. (2014). Enrichment of statistical power for genome-wide association studies. *BMC Biol.* 12, 73. doi:10.1186/s12915-014-0073-5

Li, N., Zhou, T., Geng, X., Jin, Y., Wang, X., Liu, S., et al. (2018). Identification of novel genes significantly affecting growth in catfish through GWAS analysis. *Mol. Genet. Genomics* 293, 587–599. doi:10.1007/s00438-017-1406-1

Li, W., Lu, X., Luan, S., Luo, K., Sui, J., and Kong, J. (2016). Heritability of body weight and resistance to ammonia in the Pacific white shrimp *Litopenaeus vannamei* juveniles. *Chin. J. Ocean. Limnol.* 34, 1025–1033. doi:10.1007/s00343-016-5034-0

Li, X. (2007). *Preliminary study on the functions of tmem66 gene*. China: Huazhong Agricultural University.

Li, Y., Zhou, F., Huang, J., Yang, L., Jiang, S., Yang, Q., et al. (2018). Transcriptome reveals involvement of immune defense, oxidative imbalance, and apoptosis in ammonia-stress response of the black tiger shrimp (*Penaeus monodon*). *Fish. Shellfish Immunol.* 83, 162–170. doi:10.1016/j.fsi.2018.09.026

Liang, C., Liu, J., Cao, F., Li, Z., and Chen, T. (2019). Transcriptomic analyses of the acute ammonia stress response in the hepatopancreas of the kuruma shrimp (*Marsupenaeus japonicus*). *Aquaculture* 513, 734328. doi:10.1016/j.aquaculture.2019.734328

Liao, P. Y., and Lee, K. H. (2010). From SNPs to functional polymorphism: The insight into biotechnology applications. *Biochem. Eng. J.* 49, 149–158. doi:10.1016/j.bej.2009.12.021

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). Gapit: Genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi:10.1093/bioinformatics/bts444

Liu, G., Han, Z., Jiang, D., Li, W., Zhang, W., Ye, K., et al. (2020). Genome-wide association study identifies loci for traits related to swim bladder in yellow drum (*Nibea albiflora*). *Aquaculture* 526, 735327. doi:10.1016/j.aquaculture.2020.735327

Liu, S., Vallejo, R. L., Palti, Y., Gao, G., Marancik, D. P., Hernandez, A. G., et al. (2015). Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Front. Genet.* 6, 298. doi:10.3389/fgene.2015.00298

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12, e1005767. doi:10.1371/journal.pgen.1005767

Lu, C., Kuang, Y., Zheng, X., Li, C., and Sun, X. (2019). Advances of molecular marker-assisted breeding for aquatic species. *J. Fish. China* 43, 36–53.

Lu, X., Kong, J., Luan, S., Dai, P., Meng, X., Cao, B., et al. (2016). Transcriptome analysis of the hepatopancreas in the pacific white shrimp (*Litopenaeus vannamei)* under acute ammonia stress. *PLOS ONE* 11, e0164396. doi:10.1371/journal.pone.0164396

Lyu, D., Yu, Y., Wang, Q., Luo, Z., Zhang, Q., Zhang, X., et al. (2021). Identification of growth-associated genes by genome-wide association study and their potential application in the breeding of pacific white shrimp (*Litopenaeus vannamei*). *Front. Genet.* 12, 611570. doi:10.3389/fgene.2021.611570

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genet. (Austin)* 157, 1819–1829. doi:10.1093/genetics/157.4.1819

Mishiro, T., Ishihara, K., Hino, S., Tsutsumi, S., Aburatani, H., Shirahige, K., et al. (2009). Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* 28, 1234–1245. doi:10.1038/emboj.2009.81

Montaldo, H. H., and Castillo-Juárez, H. (2017). Response to strict within-family selection with special reference to aquaculture. *Aquac. Res.* 48, 5175–5178. doi:10.1111/are.13251

Ng, P. C., and Henikoff, S. (2006). Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.* 7, 61–80. doi:10.1146/annurev.genom.7.080505.115630

Qian, Z., Li, X., Xin, J., and Liu, X. (2013). PCR-SSCP Polymorphism of CTSL gene and its correlation with growth traits of *Litopenaeus vannamei* and the different mRNA expressions of CTSL. *Acta Oceanol. Sin.* 35 (06), 121–127.

Ribaut, J., and Hoisington, D. (1998). Marker-assisted selection: New tools and strategies. *Trends Plant Sci.* 3, 236–239.

Romano, N., and Zeng, C. (2013). Toxic effects of ammonia, nitrite, and nitrate to decapod Crustaceans: A review on factors influencing their toxicity, physiological consequences, and coping mechanisms. *Rev. Fish. Sci.* 21, 1–21. doi:10.1080/10641262.2012.753404

Sonia, F., Fernando, U., Alba, L., Jennifer, S., Amaia, E., Sandra, G. E., et al. (2020). OZF is a Claspin-interacting protein essential to maintain the replication fork progression rate under replication stress. *FASEB J.* 34, 6907–6919. doi:10.1096/fj.201901926R

Spencer, C. C., Su, Z., Donnelly, P., and Marchini, J. (2009). Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet.* 5, e1000477. doi:10.1371/journal.pgen.1000477

Sun, K. (2021). *Genetic parameter evaluation and genome wide association analysis of important economic traits in Litopenaeus Vannamei*. Shanghai: Shanghai Ocean University.

Teng, S., Michonova-Alexova, E., and Alexov, E. (2008). Approaches and resources for prediction of the effects of non-synonymous single nucleotide polymorphism on protein function and interactions. *Curr. Pharm. Biotechnol.* 9, 123–133. doi:10.2174/138920108783955164

Wagschal, A., Najafi-Shoushtari, S. H., Wang, L., Goedeke, L., Sinha, S., DeLemos, A. S., et al. (2015). Genome-wide identification of microRNAs regulating cholesterol and triglyceride homeostasis. *Nat. Med.* 21, 1290–1297. doi:10.1038/nm.3980

Wan, L., Dong, L., Xiao, S., Han, Z., Wang, X., and Wang, Z. (2018). Genome wide association study for economic traits in the large yellow croaker with different numbers of extreme phenotypes. *J. Genet.* 97, 887–895. doi:10.1007/s12041-018-0973-1

Wang, Q. (2017). *Genome-wide association study and genomic selection of growth and disease resistance traits in Litopenaeus vannamei*. Qingdao: Institute of Oceanology, Chinese Academy of Sciences.

Wang, Q. (2013). *Principles and practice of breeding in aquatic organisms*. Beijing: Science Press.

Wang, Q., Tian, F., Pan, Y., Buckler, E. S., Zhang, Z., and Li, Y. (2014). A SUPER powerful method for genome wide association study. *PloS one* 9, e107684. doi:10.1371/journal.pone.0107684

Wang, Q., Yu, Y., Zhang, Q., Zhang, X., Huang, H., Xiang, J., et al. (2019a). Evaluation on the genomic selection in *Litopenaeus vannamei* for the resistance against Vibrio parahaemolyticus. *Int. J. Biol. Macromol.* 505, 212–225. doi:10.1016/j.ijbiomac.2019.05.101

Wang, Q., Yu, Y., Zhang, Q., Zhang, X., Yuan, J., Huang, H., et al. (2019b). A novel candidate gene associated with body weight in the pacific white shrimp *Litopenaeus vannamei*. *Front. Genet.* 10, 520. doi:10.3389/fgene.2019.00520

Wang, Z., and Moult, J. (2001). SNPs, protein structure, and disease. *Hum. Mutat.* 17, 263–270. doi:10.1002/humu.22

Wen, Y., Zhang, H., Ni, Y., Huang, B., Zhang, J., Feng, J., et al. (2018). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* 19, 700–712. doi:10.1093/bib/bbw145

Xiao, J., Li, Q., Tu, J., Chen, X., Chen, X., Liu, Q., et al. (2019). Stress response and tolerance mechanisms of ammonia exposure based on transcriptomics and metabolomics in *Litopenaeus vannamei*. *Ecotoxicol. Environ. Saf.* 180, 491–500. doi:10.1016/j.ecoenv.2019.05.029

Yang, J., Jiang, H., Yeh, C. T., Yu, J., Jeddeloh, J. A., Nettleton, D., et al. (2015). Extreme-phenotype genome-wide association study (XP-GWAS): A method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. *Plant J.* 84, 587–596. doi:10.1111/tpj.13029

Yu, Y. (2014). *Development of molecular markers and their applications in selective breeding of the Pacific white shrimp,* Litopeneaus vannamei. Qingdao: Institute of Oceanology, Chinese Academy of Sciences.

Yu, Y., Wang, Q., Zhang, Q., Luo, Z., Wang, Y., Zhang, X., et al. (2019). Genome scan for genomic regions and genes associated with growth trait in pacific white shrimp *Litopeneaus vannamei*. *Mar. Biotechnol.* 21, 374–383. doi:10.1007/s10126-019-09887-w

Yuan, R., Hu, Z., Liu, J., and Zhang, J. (2018). Genetic parameters for growth-related traits and survival in pacific white shrimp, *Litopenaeus vannamei* under conditions of high ammonia- N concentrations. *Turk. J. Fish. Aquat.* 18, 37–47. doi:10.4194/1303-2712-v18_1_05

Zeng, D., Yang, C., Li, Q., Zhu, W., Chen, X., Peng, M., et al. (2020). Identification of a quantitative trait loci (QTL) associated with ammonia tolerance in the Pacific white shrimp (*Litopenaeus vannamei*). *BMC Genomics* 21, 857. doi:10.1186/s12864-020-07254-x

Zhang, J., Cao, F., Liu, J., Yuan, R., and Hu, Z. (2016). Genetic parameters for growth and hypoxic tolerance traits in pacific white shrimp *Litopenaeus vannamei* at different ages. *N. Am. J. Aquacult.* 79, 75–83. doi:10.1080/15222055.2016.1194923

Zhang, X., Pan, L., Wei, C., Tong, R., Li, Y., Ding, M., et al. (2020). Crustacean hyperglycemic hormone (CHH) regulates the ammonia excretion and metabolism in white shrimp, *Litopenaeus vannamei* under ammonia-N stress. *Sci. Total Environ.* 723, 138128. doi:10.1016/j.scitotenv.2020.138128

Zhang, Z., Ersoz, E., Lai, C., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi:10.1038/ng.546

Zhou, Z., Chen, L., Dong, C., Peng, W., Kong, S., Sun, J., et al. (2018). Genome-scale Association study of abnormal scale pattern in yellow river carp identified previously known causative gene in European mirror carp. *Mar. Biotechnol.* 20, 573–583. doi:10.1007/s10126-018-9827-3

Zou, S., and Jiang, X. (2008). Retracted: Gene duplication and functional evolution of hox genes in fishes. *J. Fish Biol.* 73, 329–354. doi:10.1111/j.1095-8649.2008.01852.x

Zuo, H., Yang, L., Zheng, J., Su, Z., Weng, S., He, J., et al. (2018). A single C4 Zinc finger-containing protein from *Litopenaeus vannamei* involved in antibacterial responses. *Fish. Shellfish Immunol.* 81, 493–501. doi:10.1016/j.fsi.2018.07.053

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 29 August 2022
DOI 10.3389/fgene.2022.936610



OPEN ACCESS

EDITED BY
You-Yi Kuang,
Heilongjiang River Fisheries Research Institute (CAS), China

REVIEWED BY
Marco Tolone,
University of Palermo, Italy
Shuming Zou,
Shanghai Ocean University, China

*CORRESPONDENCE
Shengjie Li,
ssjjli@163.com

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

RECEIVED 05 May 2022
ACCEPTED 21 July 2022
PUBLISHED 29 August 2022

CITATION
Du J, Li S, Shao J, Song H, Jiang P, Lei C, Bai J and Han L (2022), Genetic diversity analysis and development of molecular markers for the identification of largemouth bass (Micropterus salmoides L.) based on whole-genome re-sequencing.
*Front. Genet.* 13:936610.
doi: 10.3389/fgene.2022.936610



# Genetic diversity analysis and development of molecular markers for the identification of largemouth bass (*Micropterus salmoides* L.) based on whole-genome re-sequencing

Jinxing Du[1], Shengjie Li[1]*, Jiaqi Shao[1], Hongmei Song[1], Peng Jiang[1], Caixia Lei[1], Junjie Bai[1] and Linqiang Han[2]

[1]Key Laboratory of Tropical and Subtropical Fishery Resources Application and Cultivation, Ministry of Agriculture and Rural Affairs, Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Guangzhou, China, [2]Guangdong Liangshi Aquatic Seed Industry Co Ltd, Foshan, China

Largemouth bass (*Micropterus salmoides* L.) is generally considered to comprise two subspecies, Florida bass (*M. floridanus*) and Northern Largemouth bass (*M. salmoides*), which have biological characteristic differences because of their geographical distribution. In this study, whole-genome re-sequencing was performed among 10 Florida and 10 Northern largemouth bass, respectively. In total, 999,793 SNPs and 227,797 InDels were finally identified, and 507,401 SNPs (50.75%) and 116,213 InDels (51.01%) were successfully mapped to annotated 18,629 genes and 14,060 genes, respectively. KEGG classification indicated that most of these genes were focused on the pathways including signal transduction, transport and catabolism, and endocrine system. Genetic diversity analysis indicated that Florida largemouth bass had higher genetic diversity than Northern largemouth bass, indicating that the germplasm quality of Northern largemouth bass had markedly reduced in China. To examine the accuracies of the identified markers, 23 SNPs and eight InDels (the insertions or deletions of more than 45 bp) were randomly selected and detected among Florida largemouth bass, Northern largemouth bass, and their F1 hybrids. The detection efficiencies of all the markers were higher than 95%; nineteen SNPs and three InDels could accurately distinguish the two subspecies and their F1 hybrids with 100% efficiencies. Moreover, the three InDel markers could clearly distinguish the two subspecies and their F1 hybrids with a PCR-based agarose gel electrophoresis. In conclusion, our study established a simple PCR-based method for the germplasm identification of largemouth bass, which will be useful in the germplasm protection, management, and hybridization breeding of largemouth bass.

KEYWORDS

**largemouth bass, genetic diversity, germplasm identification, SNP, INDEL**

## Introduction

Largemouth bass (*Micropterus salmoides* L.) is native to freshwater lakes and rivers in North America. It is considered to comprise two subspecies (Bailey and Hubbs, 1949), the Florida largemouth bass (*M. floridanus*, FB) distributed in the Florida peninsula, and the Northern largemouth bass (*M. salmoides*, NB) distributed in most central and eastern parts of America, northeast Mexico, and southeast areas of Canada (Maceina and Murphy, 1992). With its worldwide introduction in the 1970s, it was introduced from Taiwan to mainland China in 1983 and has become an economically cultured fish in China with an output of 620,000 tons in 2020. Our teams collected mainly largemouth bass populations in different areas in China and re-imported the wild FB and NB populations from America in 2010 (Cai et al., 2012). Through morphological traits, microsatellite molecular markers (Fan et al., 2009), and DNA fingerprinting (Fan et al., 2012), we confirmed that the cultured largemouth bass in China belongs to the Northern subspecies. In the past two decades, two new varieties of largemouth bass have been selectively bred by our team based on the NB population: "Youlu No.1" bred in 2011 and "Youlu No.3" bred in 2019 (Bai and Li, 2019). However, the genetic diversity of cultured largemouth bass has been markedly reduced during the breeding process (Bai et al., 2008; Fan et al., 2019).

Hybridization, as one of the important methods for fish breeding, can effectively transfer good parental traits and increase the genetic variation of offspring (Hulata 2001; Lou 2007). Currently, several hybridization studies and the comparison of the genetic structures of the hybrids and their parents have been reported in largemouth bass. In the reciprocal cross of cultured NB and imported FB in 2010, the observed heterozygosity (*Ho*), expected heterozygosity (*He*), and polymorphic information content (PIC) were all highest in FB ♀ × NB ♂, which were 0.849, 0.639, and 0.571, respectively (Cai et al., 2012). In the reciprocal cross of "Youlu No.1" and NB imported in 2010, the *Ho*, *He*, and PIC values were all highest in "Youlu No.1" ♀ × NB ♂, which was 0.729, 0.553, and 0.454, respectively (Zhou et al., 2020). Recently, in the genetic diversity of F1 progeny on cross species of FB, NB, and "Youlu No.3", the *Ho*, *He*, and PIC were all highest in "Youlu No.3" ♀ × FB ♂, which was 0.559, 0.600, and 0.521, respectively (Wang et al., 2020). These results indicated that hybridization has an obvious effect on the improvement of genetic diversity in largemouth bass.

Accurately distinguishing the hybrids and their parents is the first step in hybridization breeding of largemouth bass. Because the morphological traits of the hybrids are midway between FB and NB, it makes classification of largemouth bass difficult based on morphology alone (Rogers et al., 2006). During the last several decades, molecular genetic markers, such as mitochondrial DNA (mtDNA) analysis with restriction fragment length polymorphism (RFLP) (Nedbal and Philipp, 1994), random amplified polymorphic DNA (RAPD) (Williams et al., 1998), specific sites of the isoenzyme (Philipp et al., 1983), and microsatellite molecular markers (Lutz-Carrillo et al., 2006; Lutz-Carrillo et al., 2008; Seyoum et al., 2013), have been applied to a variety of genetic studies of largemouth bass (Austin et al., 2012; Hargrove et al., 2019). With the development of sequencing technology, the third generation of DNA markers is based on single nucleotide polymorphisms (SNPs) and insertion/deletion (InDel), have been widely used in the germplasm identification in plants and animals (Zhang et al., 2017; Guo. et al., 2019; Thongda et al., 2019; Zhao et al., 2019). Currently, gene-linked SNP markers have been discovered (Li et al., 2015) and applied to the classification of largemouth bass (Hargrove et al., 2019), while the InDel markers are less reported. Compared with the requirement of special equipment systems for SNP detection, the detection of InDel technology is user-friendly and low-cost (Guo et al., 2019), which has high application values in the germplasm classification of largemouth bass.

The obtainment of the largemouth bass reference genome (GenBank: GCA_019677235.1) provides an advantage for the discovery of SNP and InDel markers between NB and FB. In this study, 10 "pure" FB and 10 "pure" NB were selected and sequenced with a whole-genome re-sequencing technique. Then, the SNP and InDel markers were screened and annotated according to the reference genome of largemouth bass. In addition, genetic diversity analysis was conducted based on the high-quality SNPs, and a series of SNP and InDel markers were randomly selected and detected among FB, NB, and their F1 hybrids. Our study provided a PCR-based method for the germplasm identification of largemouth bass, which would be useful in the protection, management, and hybridization breeding of largemouth bass.

## Materials and methods

### Sample collection

Largemouth bass were collected from the Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences (Guangzhou, China). FB was originally imported from America in 2010, and NB was the new variety of largemouth bass "Youlu No.3" bred in 2019. Identifications of these fish were conducted based on a subset of microsatellite primers from Malloy et al. (2000) and Lutz-Carrillo et al. (2008). Then, the fin tissues used for whole-genome re-sequencing were collected from 10 FB (N1-N10) and 10 NB (F1-F10), respectively. All the fin tissues were stored at −80°C until DNA extraction.

### DNA extraction and Re-sequencing

For each sample, genomic DNA was extracted from 0.1 g fin tissue using a modified CTAB method (Stewart and Via, 1993).

The extracted DNA samples were quantified using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, United States) and their quality was assessed through 0.8% agarose gel electrophoresis. Library construction for re-sequencing was prepared with 1.0 μg of starting total DNA and processed using the VAHTS universal DNA Library Prep Kit for MGI (Vazyme, Nanjing, China) following the manufacturer's recommendations. Paired-end libraries with an insertion size of 350 bp were constructed for 10 NB and 10 FB, respectively, and index codes were added to attribute sequences to each sample. The library quantification and size were measured using the Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, United States) and the Bioanalyzer 2,100 system (Agilent Technologies, CA, United States). Subsequently, sequencing was performed on an MGI-SEQ 2000 platform by Frasergen Bioinformatics Co., Ltd. (Wuhan, China).

## Detection and annotation of SNP and InDels

After removing adapter sequences and low-quality reads, the clean reads were further rechecked for quality based on the following criteria: consecutive bases on the ends with base quality <20, read length <50 bp, and the singletons were also removed. High-quality sequences were aligned and mapped to the reference genome of largemouth bass using a BWA program (v0.7.17) (Li and Durbin, 2009) with default settings. The sequencing depth and coverage compared to the reference genome were calculated based on the alignment results using SAMtool software (Li et al., 2009).

SNPs and InDels were called using the Genome Analysis Toolkit (GATK) Haplotype Caller (McKenna et al., 2010). To reduce the error rate of calling variations, SNPs were filtered by variant filtration tools with the following threshold: QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < −12.5, and ReadPosRankSum < −8.0. Meanwhile, InDels were filtered by GATK with the recommended threshold: QD < 2.0, FS > 200.0, and ReadPosRankSum < −20.0. The mutational positions and genomic regions, were assessed using ANNOVAR (v2019). The genome-wide distribution of mutations was plotted using Circos software (v0.69). Finally, the functional annotation of the genes with mutations were predicted by Nr, SwissProt, GO, KOG, and KEGG databases with DIAMOND (ver.0.9.22.123) software.

## Genetic diversity analysis

All the SNPs that passed filtering in the population were used to construct a phylogenetic tree using TreeBeST (http://treesoft.sourceforge.net/treebest.shtml/) (Vilella et al., 2009). The final phylogenetic tree was plotted using iTOL (http:/itol.embl.de). The principal component analysis was performed using PLINK v1.07 software with default parameters (Yang et al., 2011). Population structure clustering was analyzed using Admixture (v1.3.0) with K setting from 2 to 10. The K value with minimum cross-validation error was chosen the best population structure (Yang et al., 2011). Linkage disequilibrium (LD) was calculated using PopLDdecay (v3.30) (Zhang et al., 2019), and the LD decay was calculated based on the squared correlation coefficient ($r^2$) values between the two SNPs and the physical distance between the two SNPs. Nucleotide diversity ($\pi$) and inbreeding coefficient (F) were calculated using VCFtolls (v.0.1.13) (Danecek et al., 2011).

## Experimental validation of SNP and InDel markers

In total, 130 largemouth bass came from three populations were used to validate the accuracies of candidate SNP and InDel markers: 1) 20 FB and 20 NB re-imported from America in 2010; 2) 30 FB progenies of the 2010 re-imported FB bred in 2020, and 30 NB progenies of the "Youlu three" bred in 2020; 3) 30 $F_1$ hybrids (NF1 -NF30) bred in 2020, which were obtained by "Youlu No.3" ♀ × re-imported FB in 2010 ♂. Total DNA was extracted from fin tissues using the standard phenol-chloroform procedure and DNA integrity was examined using agarose gel electrophoresis and quantity was determined using an Agilent 2,100 Bioanalyzer (Agilent, Shanghai, China).

The 30 NB, 30 FB, and 30 NF were used for the SNP detection. The polymorphisms of 23 randomly selected SNPs were detected with the Snapshot method. Briefly, primers were designed according to the largemouth bass genome sequences of these SNPs, respectively (Supplementary Table S1). Then, multiplex PCR amplification was performed with 1 μl DNA, 1 μl forward primer, and 1 μl reverse primer (10 pM), 7.5 μl Premix Taq™ (Takara, #RR901A), and 4.5 μl $ddH_2O$. The PCR procedure was as followed: 94°C for 3 min, followed by 35 cycles of 94°C for 15 s, 58°C for 15 s, 72°C for 30 s, and finally, 72°C for 4 min. Then, the PCR production was purified with ExoI and FastAP. The reaction mixture was 3 μl PCR production, 0.2 μl ExoI, 0.8 μl FastAP, 0.7 μl ExoI buffer, and 3.3 μl $ddH_2O$, respectively. The PCR procedure was as followed: 37°C for 15 min and 80°C for 15 min. Then, the extension reaction was performed with 2 μl purified PCR production, 1 μl Snapshot Ready ReactionMix, 1 μl extension primer (10 pM), and 3 μl $ddH_2O$. The procedure was as follows: 96°C for 1 min, followed by 30 cycles of 96°C for 10 s, 52°C for 5 s, and 60°C for 30 s. Then, 10 μl Hidi formamide (Applied Biosystems, Foster City, United States) was added into 1 μl extension production and treated in the ice. Then, the detection of DNA polymorphisms was sequenced with an ABI3730XL Sequencer (Applied Biosystems, Foster City, CA).

To develop the InDel markers for the germplasm identification of largemouth bass, eight InDels were

TABLE 1 Summary of the re-sequencing results of NB and FB.

| Basic information | NB | FB |
|---|---|---|
| Total clean reads | 619,675,540 | 633,811,124 |
| Total clean base(bp) | 92,025,014,273 | 94,176,077,761 |
| Length | 150 | 150 |
| Average Q20 | 97.52 | 97.81 |
| Average properly mapped (%) | 95.37% | 94.03% |
| Average singletons mapped (%) | 0.09% | 0.17% |
| Average map ratio (%) | 99.51 | 99.67 |
| Average depth (x) | 10.31 | 10.40 |
| Average cover ratio (%) | 97.33 | 95.57 |
| Average GC content (%) | 39.99 | 40.00 |

randomly selected from the identified InDels, with a size number >45 bp of the insertions or deletions bases in each of the InDel makers. These InDel markers were detected in the aforementioned NB, FB, and NF individuals. qPCR was performed in a 20 μl reaction mixture including 10 μl of Premix Taq™ (Takara, #RR901A), 0.5 μl of each primer (10 μm), 1 μl of DNA, and 8 μl of $ddH_2O$. The PCR procedure was as follows: 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 58°C for 30 s, 72°C for 30 s, and finally, 72°C for 10 min. Then, electrophoresis on a 2.0% agarose gel was used to detect the sizes of PCR amplification products.

# Results

## Analysis of genome re-sequencing data

A total of 619,675,540 and 633,811,124 paired-end clean reads were generated from 10 NB to 10 FB, respectively, which had an average coverage depth of approximately 10× to largemouth bass reference genome (*Micropterus salmoides*). The overall mapping rate was 97.33% for NB and 95.57% for FB, respectively, with an average of 96.45%. Moreover, for NB and FB, 95.37 and 94.03% paired-end reads, and 0.09 and 0.17% single-end reads were mapped to the reference chromosomes of the largemouth bass genome, indicating the high quality of the sequencing data. In addition, nucleotide statistics on the assembled scaffolds showed that the GC content is nearly 40% (Table 1).

## Genomic distribution and annotation of SNPs and InDels in largemouth bass

In total, 999,793 filtered SNPs were finally identified between NB and FB. These SNPs were disrupted across all the 23 chromosomes (Chr), and varied from 10,398 on Chr10 to 77,288 on Chr1 (Figure 1A). The densities of SNPs in the largemouth bass genome were estimated at 0.81 per kilobase (kb) each. A total of 507,401 SNPs (50.75%) were successfully mapped to the genome sequences of annotated 18,630 genes (a total of 23,901 genes in the genome) (Supplementary Table S2). The functional characterization of genes with the polymorphic SNPs was disrupted across 23 chromosomes of largemouth bass, which varied from 1,053 on Chr2 to 463 on Chr12 (Figure 1B).

In total, 227,797 filtered InDels were finally identified between NB and FB. These InDels varied from 2,669 on Chr12 to 16,845 on Chr1 (Figure 1A). The densities of InDel in the largemouth bass genome were estimated at 3.82 per kilobase (kb) each. The majority of InDels were small and ranged from 1 to 3 bp (70.51%), and InDels longer than 50 bp had the smallest proportion (0.91%) (Figure 2). A total of 16,213 InDels (51.01%) were successfully mapped to the genome sequences of the annotated 14,061 genes (Supplementary Table S3), which varied from 834 on Chr1 to 266 on Chr12 (Figure 1B).

KEGG classification suggested that these SNPs and InDel markers are widely involved in transport and catabolism, and the cellular community of cellular processes; signal transduction, signaling molecules, and interaction of environmental information processes; folding, sorting and degradation, and transcription of genetic information processes; carbohydrate metabolism and lipid metabolism of metabolism; immune system and endocrine system of organismal systems (Figures 3A,B, Supplementary Tables S4, S5). Specially, signal transduction process related genes consisted of most polymorphic SNPs and InDels.

## Genetic diversity analysis of FB and NB

To explore the relationships between the 10 NB and 10 FB, a neighbor-joining phylogenetic tree was constructed using all the filtered SNPs. The phylogenetic tree classified the 20 LMB into two groups: the 10 FB were clustered into one group and the 10 NB clustered into another group (Figure 4A). A population genetic structure analysis was performed based on the high-quality SNPs. We employed 5-fold cross-validation to infer the number of ancestral populations (Figure 4B; Supplementary Figure S1). At K = 2, NB showed a strong genetic differentiation from FB. At K = 3, the population structure of FB was complex, especially when the K value was higher, suggesting a higher genetic diversity in FB (Figure 4B).

Principal component analysis (PCA) was used to further confirm the relationship between the 10 NB and 10 FB. As shown in the principal component plot for the first two principal components, which accounted for 60.24% of the total variation observed in 20 largemouth bass. PC1 explained 55.71% of the overall variation and separated 20 largemouth

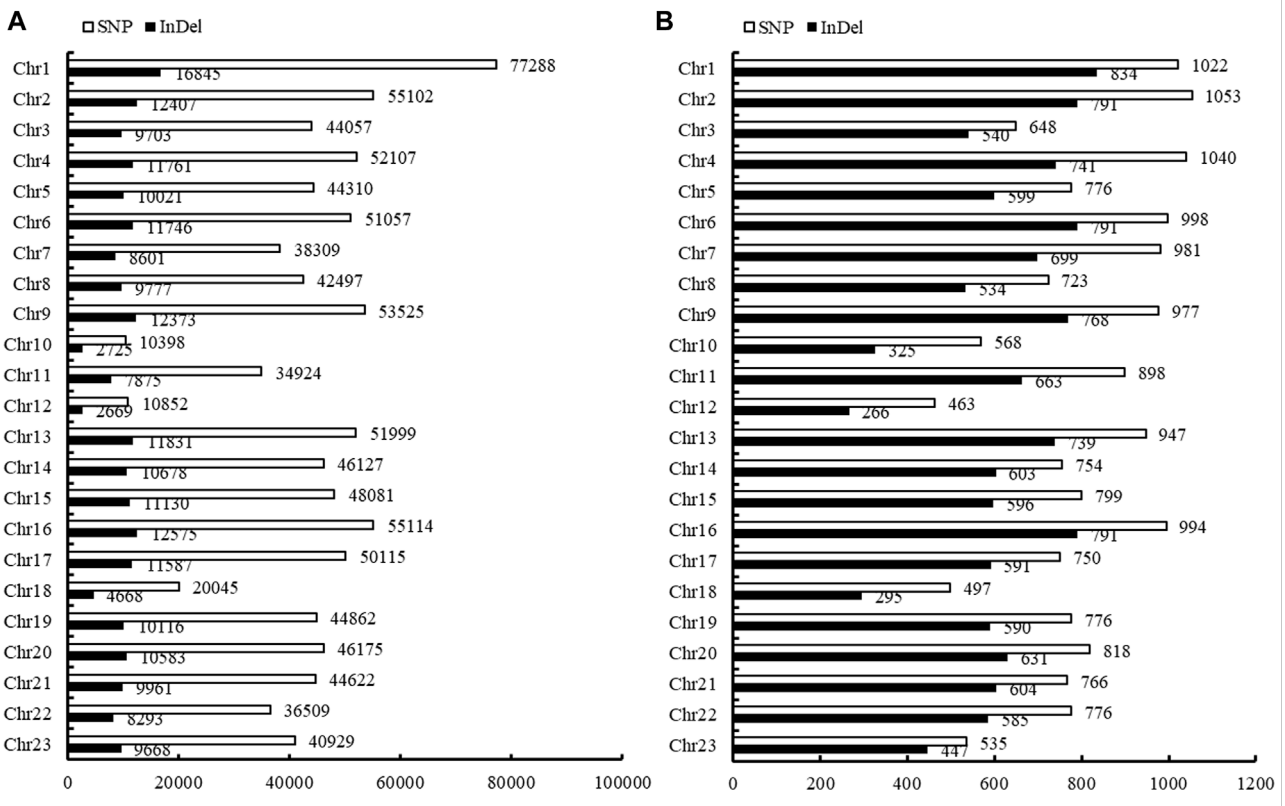



FIGURE 1
Disruption of SNP and InDels in the genome and genes of Largemouth bass. **(A)** Disruption of SNP and InDels in each chromosome of the largemouth bass. **(B)** Annotation of the SNPs and InDels in the corresponding genes of the largemouth bass.

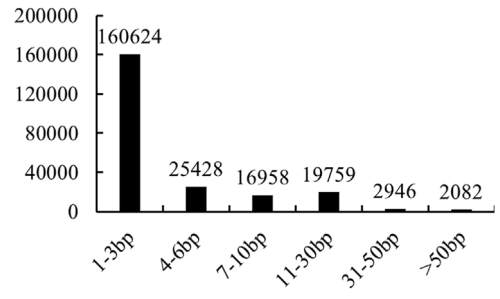



FIGURE 2
Summary of InDels between NB and FB.

bass into two groups named NB and FB, and PC2 explained 4.53% of the overall variation (Figure 4C). The LD decay rate in FB was faster than that in NB (Figure 4D). Moreover, the nucleotide diversity ($\pi$) of FB was higher than that of NB, while the inbreeding coefficient (F) of FB was lower than that of NB. Taken together, our results indicated that the NB population experienced a longer breeding process.

## Validation of SNP and InDel polymorphisms in FB, NB, and their F1 progenies

To test the efficiency of the identified SNPs, a total of 23 randomly selected SNPs were conducted in 30 NB, 30 FB, and their 30 F1 progenies (NF), with one on each chromosome. The results indicated that all the 23 SNPs performed good polymorphisms and 19 SNPs (82.60%) could distinguish these fish with 100% efficiencies (Table 2; Supplementary Table S6). The accuracies of the other four SNPs ranged from 96.67 to 97.78%, respectively.

In order to develop a PCR-based method for the germplasm identification of largemouth bass, a total of eight InDel markers were selected randomly from the aforementioned Indels, which were longer than 45 bp (Table 3). After PCR amplification with specific primers among 50 NB, 50 FB, and 30 NF, the PCR production were detected with 2% agarose gel electrophoresis, and only three InDel markers (ID1, ID5, and ID8) could accurately identify the NB, FB, and NF with 100% accuracies, with one different length stripe for NB and FB, respectively, and two stripes for the hybrids (Figure 5; Supplementary Figure S2).

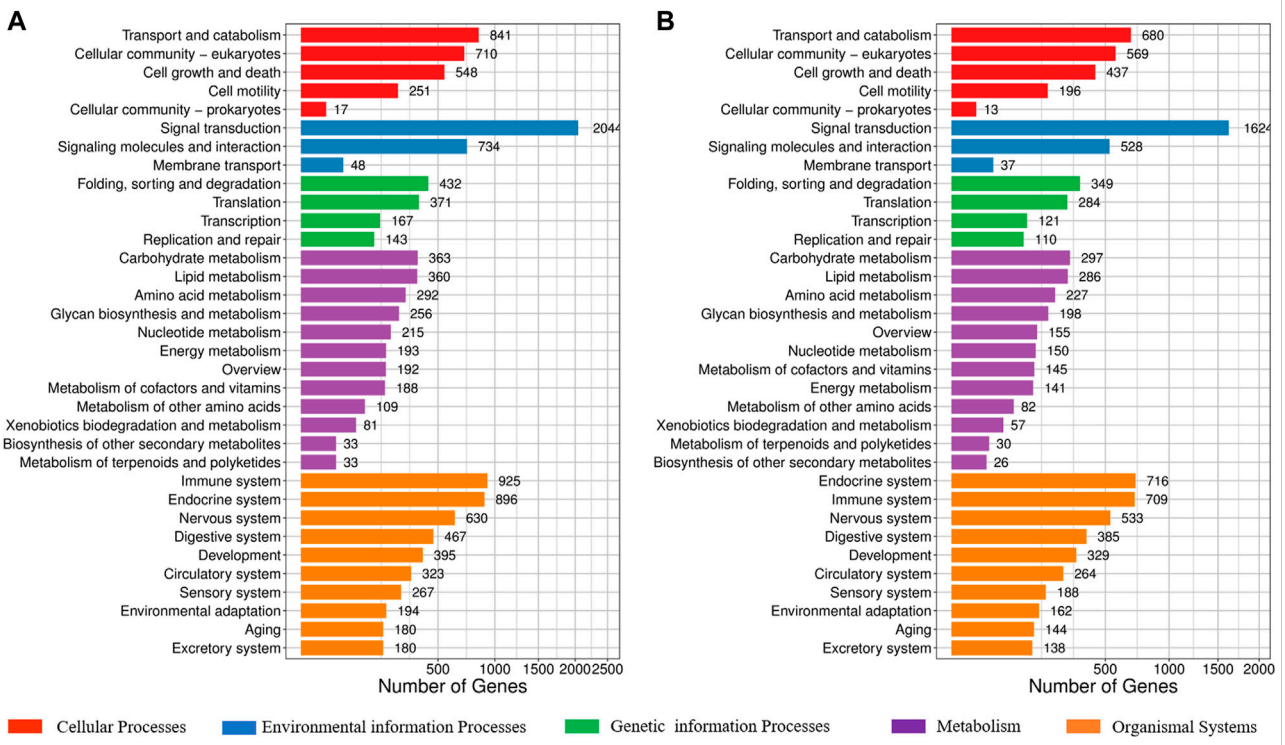



FIGURE 3
KEGG classification of SNPs and InDels in the potential biology process in the largemouth bass. **(A)** KEGG classification of SNPs in the potential biology process. **(B)** KEGG classification of InDels in the potential biology process.

In addition, the efficiencies of other five InDel markers ranged from 96.15 to 97.69%, respectively.

# Discussion

## Genome variation between FB and NB largemouth bass

To better understand the genetic variation in largemouth bass subspecies, 10 NB and 10 FB were selected for whole-genome re-sequencing. With the guidance of the largemouth bass reference genome, 999,793 SNPs were finally identified, which was much more than that in the previous studies, without the largemouth bass reference genome. For example, 3,674 SNPs were discovered through the transcriptomes analysis between NB and FB (Li et al., 2015); and 58,450 genome-wide SNPs for FB were generated using a cost-effective genotyping-by-sequencing method (Zhao et al., 2018). Moreover, 19 of 23 randomly selected SNPs (82.60%) could effectively distinguish the 30 NB, 30 FB, and 30 F1 progenies with 100% efficiencies, indicating a high quality of the sequencing data.

Similar to SNPs, InDels are widely distributed in the genome and have been used in various biological process analyses in animals and plants (Zhang et al., 2017; Guo et al., 2019; Jain et al., 2019). However, few studies have been reported about the InDels in largemouth bass. Only a 66-bp deletion in the growth hormone releasing hormone gene was reported to be related to livability during embryonic development (Ma et al., 2014), and two InDels were identified as sex-specific markers in largemouth bass (Du et al., 2021). In this study, InDels between NB and FB were analyzed for the first time. The number of 1-3bp InDels was 160,624, making up 70.51% of all the InDels in the largemouth bass genome, and the number of InDels longer than 30 bp was 5,028. Together with the identified SNPs, our results provided useful information for the functional genomics studies for largemouth bass.

## Genetic structure analysis between NB and FB

In the past two decades, we analyzed the genetic structure of collected NB largemouth bass populations with microsatellite DNA markers, including the imported NB population in 2010, the new variety “Youlu No.1” bred in 2011, and the offspring of “Youlu No.1” bred in 2016. The *Ho* values of the three populations were 0.519, 0.480, and 0.4206, respectively; the *He*

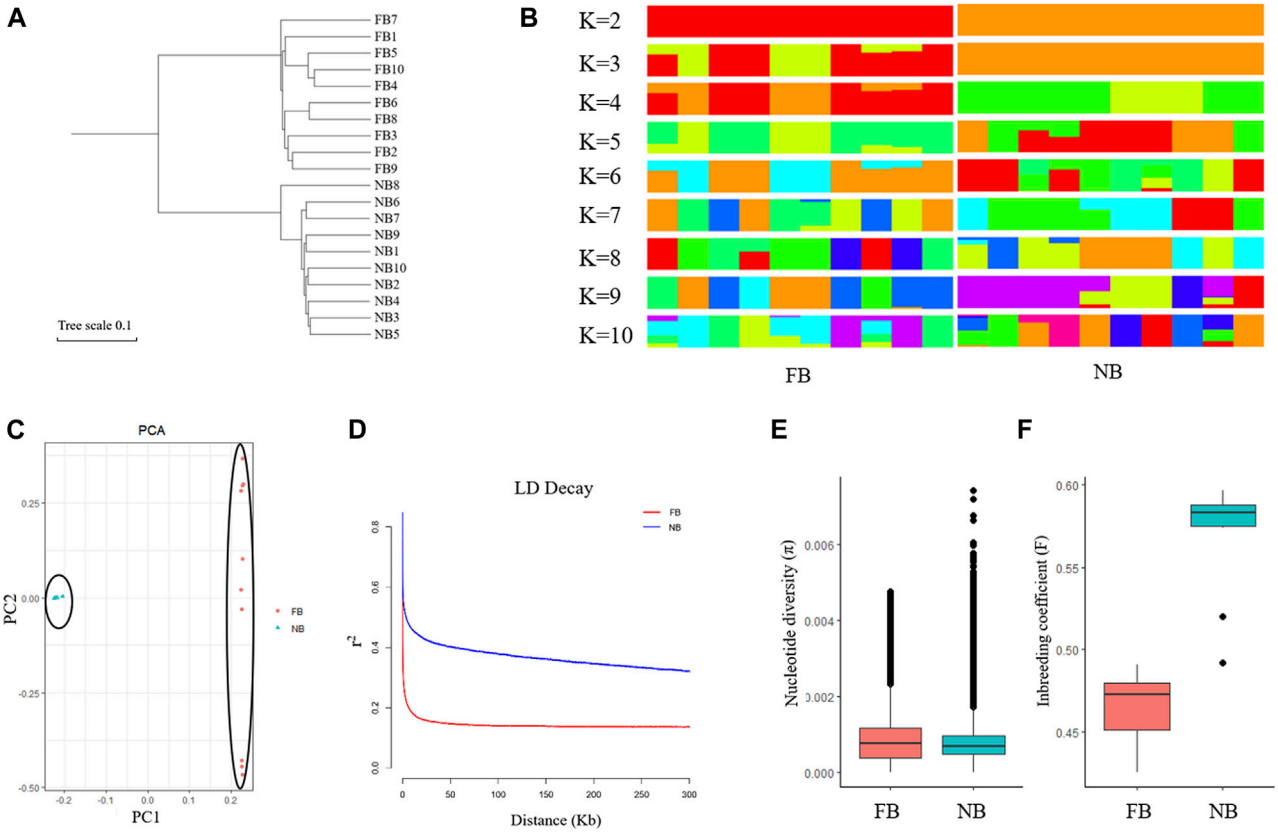



FIGURE 4
Genetic diversity analysis among 10 NB and 10 FB. **(A)** The phylogenetic tree analysis of 10 NB and 10 FB. **(B)** Population genetic structure analysis of 10 NB and 10 FB. **(C)** Principal component analysis (PCA) of10 NB and 10 FB. **(D)** Linkage disequilibrium (LD) analysis of 10 NB and 10 FB. **(E)** Nucleotide diversity ($\pi$) of 10 NB and 10 FB. **(F)** Inbreeding coefficient **(F)** of 10 NB and 10 FB.

values of the three populations were 0.491, 0.454, 0.3916, respectively; the PIC values of the three populations were 0.407, 0.412, and 0.3257, respectively (Fan et al., 2019; Zhou et al., 2020). In this study, the genetic structure of NB and FB was first analyzed on the whole-genome level based on the SNPs. The population structure, PCA analysis, and LD decay rate indicated a lower genetic diversity of the NB. The nucleotide diversity ($\pi$) of NB was lower than those in FB, corresponding to the lower genetic diversity of the NB. In addition, the inbreeding coefficient (F) of NB was higher than FB, indicating the long-time inbreeding resulted in low genetic heterogeneities in NB.

## Application of SNP and InDel markers for breeding of largemouth bass

Because of their geographical distribution differences, it was reported that NB is more tolerant to low temperature, ammonia nitrogen, and low oxygen, while FB is more resistant to high temperatures and has a stronger stress response (Philipp et al., 1983; Fields et al., 1987; Carmichael et al., 1988; Williamson and Carmichael, 1990; Khosa et al., 2020). Abiotic stresses (e.g., heat, chilling, nutritional imbalance, and water quality) are one of the major factors which restrict the growth and development of largemouth bass (Li et al., 2020; White et al., 2020; Jia et al., 2022). Especially the high temperatures in the summer and autumn seasons, which commonly result in reduced food intake, declined growth rates and disease resistance of largemouth bass (Lin et al., 2022). In this study, through the annotation and KEGG classification of these identified SNPs and InDel markers, signal transduction, immune system, and endocrine system were found to be the top three pathways. Their results provided valuable information about the genetic mechanism of heat tolerance and stress response between the two subspecies of largemouth bass.

Through gene combination of two or more species, populations, or varieties, which have different genetic bases, hybridization alters the genetic structure of the offspring and thus increases their heterozygosity levels and genetic variation (Lippman and Zamir 2007). In this study, we detected 23 candidate SNPs and eight InDel markers among FB, NB, and their hybrids. It was noticed that most of these markers were homozygous in parents and heterozygous in the

TABLE 2 Genotypes and annotation of 23 randomly selected SNPs in 30 NB, 30 FB, and 30 NF individuals, respectively.

| Number/disruption | Genotype | | | Annotated gene or related region |
|---|---|---|---|---|
| | NB (number) | FB(number) | NF(number) | |
| SNP1/Chr1 | AA (30) | GG (30) | AG (30) | Intergenic region |
| SNP2/Chr2 | TT (30) | AA (30) | AT (30) | Son of seven less homolog 1 |
| SNP3/Chr3 | GG (30) | AA (30) | AG (30) | Intergenic region |
| SNP4/Chr4 | GG (30) | AA (30) | AG (30) | Leucine-rich repeat N-terminal domain |
| SNP5/Chr5 | TT (30) | GG (30) | GT (30) | 5-Formyltetrahydrofolate cyclo-ligase |
| SNP6/Chr6 | AA (30) | CC (30) | AC (30) | Ubiquitin-conjugating enzyme E2 L3b |
| SNP7/Chr7 | TT (30) | CC (27) | GT (30) | Epidermal growth factor-like domain |
| SNP8/Chr8 | TT (28) | CC (30) | GT (30) | Nuclear apoptosis inducing factor 1 |
| SNP9/Chr9 | CC (30) | AA (30) | AC (30) | Intergenic region |
| SNP10/Chr10 | GG (30) | TT (30) | GT (30) | Prohibition |
| SNP11/Chr11 | CC (30) | TT (30) | CT (30) | Synaptotagmin XIa |
| SNP12/Chr12 | CC (30) | TT (30) | CT (30) | Intergenic region |
| SNP13/Chr13 | GG (30) | AA (30) | AG (30) | Intergenic region |
| SNP14/Chr14 | GG (30) | AA (30) | AG (30) | Solute carrier family 35 member |
| SNP15/Chr15 | AA (30) | TT (30) | AT (30) | Peptide methionine sulfoxide reductase MsrA |
| SNP16/Chr16 | TT (30) | AA (30) | AT (30) | Intergenic region |
| SNP17/Chr17 | TT (30) | AA (30) | AT (30) | Intergenic region |
| SNP18/Chr18 | GG (30) | TT (30) | TT (30) | Intergenic region |
| SNP19/Chr19 | CC (30) | AA (30) | AC (30) | Microtubule-actin cross-linking factor 1 |
| SNP20/Chr20 | TT (28) | CC (30) | GT (30) | Holocarboxylase synthetase |
| SNP21/Chr21 | AA (30) | TT (30) | AT (30) | Intergenic region |
| SNP22/Chr22 | AA (30) | GG (30) | AG (30) | Ankyrin 3b |
| SNP23/Chr23 | AA (28) | TT (30) | AT (30) | Podocalyxin-like protein |

TABLE 3 Primers and annotation of eight selected InDels in largemouth bass.

| Indel location | Insertion/deletion size | Primer (5′-3′) | Annotated gene or related region |
|---|---|---|---|
| ID1/Chr17 | +100 | F: ACATTCAGCCCTCTTGACCG | Intergenic region |
| | | R: GACACGGGGAGATCATGCAA | |
| ID2/Chr2 | +90 | F: CCTTTGTTAACCTGCCCCCT | Intergenic region |
| | | R: GTAGTCATGGGACCATCCCC | |
| ID3/Chr3 | +46 | F: GCATCGTTTCCACAGGTGTC | Intergenic region |
| | | R: GCAGCTTCCAATGCAACTGTA | |
| ID4/Chr6 | +87 | F: TCACGCCACATCCAGGTAAG | Intergenic region |
| | | R: TGCCATAGGTAACTCCCCAGT | |
| ID5/Chr20 | +83 | F: CGTGTCAGCTAACTACACCTGA | Intergenic region |
| | | R: ATACTGCCCCGCAAAGGAAA | |
| ID6/Chr5 | −84 | F: GTCAACCGGTGAACACAACG | Neural-cadherin-like |
| | | R: ACGTTATCAGCACTGTGCCA | |
| ID7/Chr17 | −51 | F: AGGGAGAAACCTCATTGGGC | Intergenic region |
| | | R: TTGCTGGCATCCTCCATAGC | |
| ID8/Chr23 | −61 | F: CACCAGCCTGCAGGTAAGAA | Intergenic region |
| | | R: CTTCCAACCACACAAGGTCAG | |

"+" and "-" mean insert and deletion in NB, in comparison with FB, respectively.

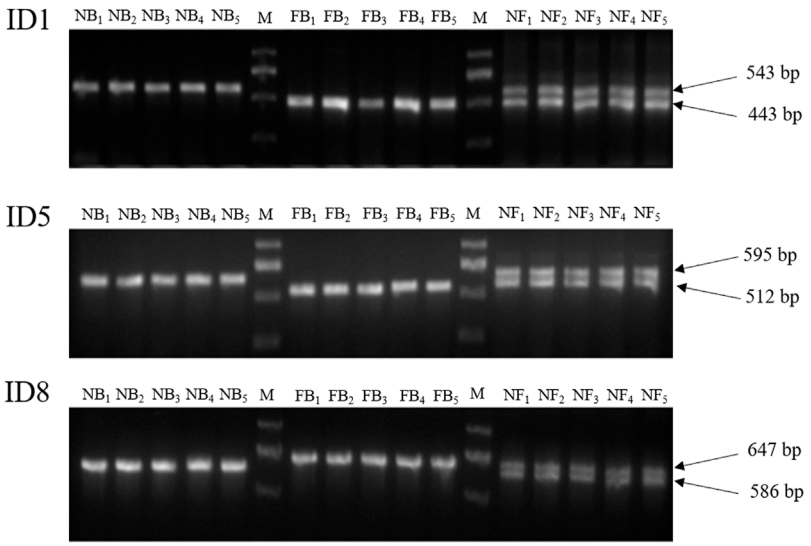



FIGURE 5
PCR amplification results of ID1, ID5, and ID8 in 5 NB, five FB, and five NF individuals, respectively. M, DNA Marker DL 2000. NB, Northern largemouth bass. FB, Florida largemouth bass. NF, the F1 hybrid of the Northern largemouth bass ♀ × Florida largemouth bass ♂.

hybrids, indicating the heterozygosity levels were increased in the hybrids. Many studies have addressed the hybridization of FB and NB, while the divergence between hybrids may be advantageous in terms of their growth performance. Indeed, most experiment results demonstrated that the NB exhibited the best growth at the age of 1 year, followed by the hybrids and then the FB (Zolczynski and Davies 1976; Williamson and Carmichael 1990; Philipp and Whitt, 1991). Our previous study also indicated that the NB exhibited better growth than the hybrids. By contrast, a previous study suggested that a hybrid of NB♂ × FB♀ grows better than the two parents (Kleinsasser et al., 1990). This difference may be related to the culture environment and methods employed, as well as the different stages of fish. In terms of heat tolerance, it was reported that both the rank order of values for critical thermal maximum and chronic thermal maximum were NB♂ × FB♀ > FB > NB♀ × FB♂ > NB (Fields et al., 1987). Therefore, it is worth attempting to produce a hybridization progeny line, which might include the advantages in both the NB and FB. Moreover, the three InDels identified in this study will provide a PCR-based method for the germplasm identification in hybridization breeding of largemouth bass.

## Conclusion

Through whole-genome resequencing of 10 NB and 10 FB, 999,793 SNPs and 227,797 InDels were finally identified, which contributed to exploring the biological characteristic differences between NB and FB. The genetic structure analysis indicated that FB had higher genetic diversity than NB, indicating that the germplasm quality of cultured largemouth bass had markedly reduced in China. In addition, three identified InDels provided a simple PCR-based method to distinguish NB, FB, and their F1-based progenies. Our study provides an effective and accurate method for the germplasm identification of largemouth bass, which will be useful in the molecular-assisted breeding of largemouth bass in the future.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA858961.

## Ethics statement

The experiments involving largemouth bass in this study were approved by the Animal Research and Ethics Committee of the Institute of Hydrobiology, Chinese Academy of Sciences.

## Author contributions

JD conceptualized the project, acquired funding, designed and conducted the experiments, analyzed the data, and wrote the manuscript. SL conceptualized the project, acquired funding,

supervised the study, and modified the manuscript. HS helped in modifying the manuscript. JS, PJ, and CL helped in collecting the samples. JB helped in designing the experiments and modified the manuscript. LH helped in providing the experimental facilities and culturing the experimental fish. All authors read and approved the final manuscript.

## Funding

This project was supported by the Central Public-Interest Scientific Institution Basal Research Fund, CAFS, with grant numbers 2021SJ-XK4 and 2022CG02, and the Science and Technology Program of Guangzhou, China, with grant number 202201010211.

## Conflict of interest

LH was employed by the company Guangdong Liangshi Aquatic Seed Industry Co., Ltd.

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

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.936610/full#supplementary-material

## References

Austin, J. D., Johnson, A., Matthews, M., Tringali, M. D., Porak, W. F., and Allen, M. S. (2012). An assessment of hatchery effects on Florida bass (*Micropterus salmoides floridanus*) microsatellite genetic diversity and sib-ship reconstruction. *Aquac. Res.* 43 (4), 628–638. doi:10.1111/j.1365-2109.2011.02873.x

Bai, J., and Li, S. (2019). *Genetic breeding and molecular marker-assisted selective breeding of largemouth bass*. 1st ed. Academic Press by Elsevier Inc,. 89–131.

Bai, J., Lutz-Carrillo, D. J., Quan, Y., and Liang, S. (2008). Taxonomic status and genetic diversity of cultured largemouth bass *Micropterus salmoides* in China. *Aquaculture* 278, 27–30. doi:10.1016/j.aquaculture.2008.03.016

Bailey, R. M., and Hubbs, C. L. (1949). The black basses (*Micropterus*) of Florida, with description of a new species. *University of Michigan Museum of Zoology Occasional Papers* 516, 1–40.

Cai, L., Bai, J., Li, S., Chen, K., Fan, J., and Ma, D. (2012). Genetic analysis of northern largemouth bass, Florida largemouth bass, and their reciprocal hybrids. *J. Fish. Sci. China* 19, 70–76. doi:10.3724/sp.j.1118.2012.00070

Carmichael, G. J., Williamson, J. H., Woodward, C. A. C., and Tomasso, J. R. (1988). Communications: Responses of northern, Florida, and hybrid largemouth bass to low temperature and low dissolved oxygen. *Progressive Fish-Culturist* 50 (4), 225–231. doi:10.1577/1548-8640(1988)050<0225:cronfa>2.3.co;2

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). 1000 genomes project analysis GroupThe variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

Du, J., Zhou, J., Li, S., Shao, J., Jiang, P., Dong, C., et al. (2021). A PCR-based method for genetic sex identification and evidence of the XX/XY sex determination system in largemouth bass (*Micropterus salmoides* L.) *Aquaculture* 545, 737220. doi:10.1016/j.aquaculture.2021.737220

Fan, J., Bai, J., Li, S., Ren, K., and Ye, X. (2012). Establishment of DNA fingerprinting and analysis on genetic structure of largemouth bass with microsatellite. *Acta hydribiol. Sin.* 36, 600–609. doi:10.3724/SP.J.1035.2012.00600

Fan, J., Bai, J., Ye, X., Li, S., and He, X. (2009). Taxonomic status of largemouth bass Micropterus Salmoides cultured in China. *J. Dalian Fish. Univ.* 24, 83–86. doi:10.3969/j.issn.1000-9957.2009.01.017

Fan, J. J., Bai, J. J., Li, S. J., Ma, D. M., and Jiang, P. (2019). Analysis on genetic diversity of three breeding populations of largemouth bass using formulated feeds. *Prog. Fish. Sci.* 40 (4), 57–64. doi:10.19663/j.issn2095-9869.20180420002

Fields, R., Lowe, S. S., Kaminski, C., Whitt, G. S., and Philipp, D. P. (1987). Critical and chronic thermal maxima of northern and Florida largemouth bass and their reciprocal F1 and F2 hybrids. *Trans. Am. Fish. Soc.* 116, 856–863. doi:10.1577/1548-8659(1987)116<856:cactmo>2.0.co;2

Guo, G., Zhang, G., Pan, B., Diao, W., Liu, J., Ge, W., et al. (2019). Development and application of InDel markers for *Capsicum spp.* based on whole-genome re-sequencing. *Sci. Rep.* 9 (1), 3691. doi:10.1038/s41598-019-40244-y

Hargrove, J. S., Weyl, O. L. F., Zhao, H., Peatman, E., and Austin, J. D. (2019). Using species-diagnostic SNPs to detail the distribution and dynamics of hybridized black bass populations in southern Africa. *Biol. Invasions* 21 (5), 1499–1509. doi:10.1007/s10530-018-01912-8

Hulata, G. (2001). Genetic manipulations in aquaculture: A review of stock improvement by classical and modern technologies. *Genetica* 111 (1-3), 155–173. doi:10.1023/a:1013776931796

Jain, A., Roorkiwal, M., Kale, S., Garg, V., Yadala, R., and Varshney, R. K. (2019). InDel markers: An extended marker resource for molecular breeding in chickpea. *PLoS One* 14 (3), e0213999. doi:10.1371/journal.pone.0213999

Jia, S. P., Wang, L., Zhang, J. M., Zhang, L., Ma, F. R., Huang, M. L., et al. (2022). Comparative study on the morphological characteristics and nutritional quality of largemouth bass (*Micropterus salmoides*) cultured in an aquaculture system using land-based container with recycling water and a traditional pond system. *Aquaculture* 549, 737721. doi:10.1016/j.aquaculture.2021.737721

Khosa, D., South, J., Cuthbert, R. N., Wasserman, R. J., and Weyl, O. L. F. (2020). Temperature regime drives differential predatory performance in Largemouth Bass and Florida Bass. *Environ. Biol. Fishes* 103 (1), 67–76. doi:10.1007/s10641-019-00933-z

Kleinsasser, L. J., Williamson, J. H., and Whiteside, B. G. (1990). Growth and catchability of northern, Florida, and F, hybrid largemouth bass in Texas ponds. *N. Am. J. Fish. Manag.* 10, 462–468. doi:10.1577/1548-8675(1990)010<0462:gaconf>2.3.co;2

Li, C., Gowan, S., Anil, A., Beck, B. H., Thongda, W., Kucuktas, H., et al. (2015). Discovery and validation of gene-linked diagnostic SNP markers for assessing hybridization between Largemouth bass (*Micropterus salmoides*) and Florida bass (*M. floridanus*). *Mol. Ecol. Resour.* 15 (2), 395–404. doi:10.1111/1755-0998.12308

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Li, X., Zheng, S., Ma, X., Cheng, K., and Wu, G. (2020). Effects of dietary starch and lipid levels on the protein retention and growth of largemouth bass (*Micropterus salmoides*). *Amino Acids* 52 (6-7), 999–1016. doi:10.1007/s00726-020-02869-6

Lin, S., Chen, Y., Zhou, W., Xue, Y., and Zhai, X. (2022). Nutritional regulation strategies for high-quality development of largemouth bass. *Feed Ind.* 43, 12–17. doi:10.13302/j.cnki.fi.2022.10.002

Lippman, Z. B., and Zamir, D. (2007). Heterosis: Revisiting the magic. *Trends Genet.* 23 (2), 60–66. doi:10.1016/j.tig.2006.12.006

Lou, Y. (2007). Close hybridization of fish and its application in aquaculture. *J. Fish. China* 31, 532–538. doi:10.3321/j.issn:1000-0615.2007.04.017

Lutz-Carrillo, D. J., Hagen, C., Dueck, L. A., and Glenn, T. C. (2008). Isolation and characterization of microsatellite loci for Florida largemouth bass, *Micropterus salmoides floridanus*, and other micropterids. *Mol. Ecol. Resour.* 8 (1), 178–184. doi:10.1111/j.1471-8286.2007.01917.x

Lutz-Carrillo, D. J., Nice, C. C., Bonner, T. H., Forstne, M. R. J., and Fries, L. T. (2006). Admixture analysis of Florida largemouth bass and northern largemouth bass using microsatellite loci. *Trans. Am. Fish. Soc.* 135, 779–791. doi:10.1577/t04-221.1

Ma, D. M., Han, L. Q., Bai, J. J., Li, S. J., Fan, J. J., Yu, L. Y., et al. (2014). A 66-bp deletion in growth hormone releasing hormone gene 5'-flanking region with largemouth bass recessive embryonic lethal. *Anim. Genet.* 45, 421–426. doi:10.1111/age.12143

Maceina, M. J., and Murphy, B. R. (1992). Stocking Florida largemouth bass outside its native range. *Trans. Am. Fish. Soc.* 121, 686–691. doi:10.1577/1548-8659-121.5.686

Malloy, T. P., Bussche, R. A., Coughlin, W. D., and Echelle, A. A. (2000). Isolation and characterization of microsatellite loci in smallmouth bass, *Micropterus dolomieu* (Teleostei : Centrarchidae), and cross-species amplification in spotted bass. *Mol. Ecol.* 9, 1946–1948. doi:10.1046/j.1365-294x.2000.01096-16.x

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110

Nedbal, M. A., and Philipp, D. P. (1994). Differentiation of mitochondrial DNA in largemouth bass. *Trans. Am. Fish. Soc.* 123, 460–468. doi:10.1577/1548-8659(1994)123<0460:domdil>2.3.co;2

Philipp, D. P., Childers, W. F., and Whitt, G. S. (1983). A biochemical genetic evaluation of the northern and Florida subspecies of largemouth bass. *Trans. Am. Fish. Soc.* 112 (1), 1–20. doi:10.1577/1548-8659(1983)112<1:abgeot>2.0.co;2

Philipp, D. P., and Whitt, G. S. (1991). Survival and growth of northern, Florida, and reciprocal F1 hybrid largemouth bass in Central Illinois. *Trans. Am. Fish. Soc.* 120, 711–722.

Rogers, M. W., Allen, M. S., and Porak, W. F. (2006). Separating genetic and environmental influences on temporal spawning distributions of largemouth bass (*Micropterus salmoides*). *Can. J. Fish. Aquat. Sci.* 63, 2391–2399. doi:10.1139/f06-122

Seyoum, S., Barthel, L. B., Tringali, M. D., Davis, M. C., Schmitt, S. L., Bellotti, P. S., et al. (2013). Isolation and characterization of eighteen microsatellite loci for the largemouth bass, *Micropterus salmoides*, and cross amplification in congeneric species. *Conserv. Genet. Resour.* 5, 697–701. doi:10.1007/s12686-013-9885-9

Stewart, C. N., and Via, L. E. (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *BioTechniques* 14, 748–750.

Thongda, W., Lewis, M., Zhao, H., Bowen, B., Lutz-Carrillo, D. J., Peoples, B. K., et al. (2019). Species-diagnostic SNP markers for the black basses (*Micropterus* spp.): A new tool for black bass conservation and management. *Conserv. Genet. Resour.* 12 (2), 319–328. doi:10.1007/s12686-019-01109-8

Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.* 19 (2), 327–335. doi:10.1101/gr.073585.107

Wang, P., Zhou, G., Chen, S., and Lu, J. (2020). Analysis of growth trait comparison and genetic diversity of F1 progeny on cross species of southern largemouth bass, northern largemouth bass and "Youlu No.3". *Mar. Fish.* 42, 403–409. doi:10.13233/j.cnki.mar.fish.2020.04.003

White, D. P., Nannini, M. A., and Wahl, D. H. (2020). Examining the effects of chronic, lake-wide elevated temperatures on behavioural expression in largemouth bass, *Micropterus salmoides*. *J. Fish. Biol.* 97 (1), 39–50. doi:10.1111/jfb.14313

Williams, D. J., Kazianis, S., and Walter, R. B. (1998). Use of random amplified polymorphic DNA (RAPD) for identification of largemouth bass subspecies and their intergrades. *Trans. Am. Fish. Soc.* 127, 825–832. doi:10.1577/1548-8659(1998)127<0825:uorapd>2.0.co;2

Williamson, J. H., and Carmichael, G. J. (1990). An aquacultural evaluation of Florida, northern, and hybrid largemouth bass, *Micropterus salmoides*. *Aquaculture* 82, 247–257. doi:10.1016/0044-8486(90)90024-h

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88 (1), 76–82. doi:10.1016/j.ajhg.2010.11.011

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35 (10), 1786–1788. doi:10.1093/bioinformatics/bty875

Zhang, T., Gu, M., Liu, Y., Lv, Y., Zhou, L., Lu, H., et al. (2017). Development of novel InDel markers and genetic diversity in *Chenopodium quinoa* through whole-genome re-sequencing. *BMC Genomics* 18 (1), 685. doi:10.1186/s12864-017-4093-8

Zhao, H., Fuller, A., Thongda, W., Mohammed, H., Abernathy, J., Beck, B., et al. (2019). SNP panel development for genetic management of wild and domesticated white bass (*Morone chrysops*). *Anim. Genet.* 50 (1), 92–96. doi:10.1111/age.12747

Zhao, H., Li, C., Hargrove, J. S., Bowen, B. R., Thongda, W., Zhang, D., et al. (2018). SNP marker panels for parentage assignment and traceability in the Florida bass (*Micropterus floridanus*). *Aquaculture* 485, 30–38. doi:10.1016/j.aquaculture.2017.11.014

Zhou, J., Li, S., Jiang, P., Sun, X., Han, L., and Bai, J. (2020). Comparison analysis of genetic diversity and growth traits among "Youlu No. 1" and their reciprocal hybrids of northern *Micropterus salmoides*. *Mar. Fish.* 42, 324–331. doi:10.13233/j.cnki.mar.fish.2020.03.008

Zolczynski, J. R. S. J., and Davies, D. W. (1976). Growth characteristics of the northern and Florida subspecies of largemouth bass and their hybrid, and a comparison of catchability between the subspecies. *Trans. Am. Fish. Soc.* 105, 240–243. doi:10.1577/1548-8659(1976)105<240:gcotna>2.0.co;2

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 16 September 2022
DOI 10.3389/fgene.2022.1007548



OPEN ACCESS

EDITED BY
Siti Nor,
University of Malaysia Terengganu, Malaysia

REVIEWED BY
Xin Qi,
Ocean University of China, China
Vinicius Farias Campos,
Federal University of Pelotas, Brazil

*CORRESPONDENCE
Takashi Sakamoto,
takashis@kaiyodai.ac.jp

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

RECEIVED 30 July 2022
ACCEPTED 25 August 2022
PUBLISHED 16 September 2022

CITATION
Hattori RS, Kumazawa K, Nakamoto M, Nakano Y, Yamaguchi T, Kitano T, Yamamoto E, Fuji K and Sakamoto T (2022), Y-specific amh allele, amhy, is the master sex-determining gene in Japanese flounder Paralichthys olivaceus.
*Front. Genet.* 13:1007548.
doi: 10.3389/fgene.2022.1007548



# Y-specific *amh* allele, *amhy*, is the master sex-determining gene in Japanese flounder *Paralichthys olivaceus*

Ricardo Shohei Hattori[1], Keiichiro Kumazawa[1], Masatoshi Nakamoto[1], Yuki Nakano[1], Toshiya Yamaguchi[2], Takeshi Kitano[3], Eiichi Yamamoto[4], Kanako Fuji[1] and Takashi Sakamoto[1]*

[1]Department of Marine Biosciences, Tokyo University of Marine Science and Technology, Tokyo, Japan, [2]Nansei Field Station, National Research and Development Agency, Japan Fisheries Research and Education Agency, Mie, Japan, [3]Department of Biological Sciences, Graduate School of Science and Technology, Kumamoto University, Kumamoto, Japan, [4]Tottori Prefectural Fisheries Experimental Station, Tottori, Japan

Japanese flounder (*Paralichthys olivaceus*) is an important marine fish species of both fisheries and aquaculture in Northeast Asia. The commercial interest for all-female progenies due to several sex-related traits has prompted basic research on the mechanisms of sex determination in this species. By conducting a linkage analysis of the sex-determining locus, we initially identified 12 microsatellite markers linked to sex in 11 scaffolds, whose localization was restricted to a specific region of linkage group 9. Sequence analysis of this region identified 181 genes based on the UniProt database annotations. Among them, the *amh* gene was considered a potential candidate for sex determination because this gene is known to have taken over the role of sex determination in many teleosts. An in-depth sequence analysis of both the coding and non-coding regions of *amh* in XX and XY individuals detected nine SNPs linked with maleness. However, because these substitutions were synonymous, the upstream and downstream regions of *amh* were also investigated and a male-specific variant with deletions in the promoter region was detected. This truncated Y-specific *amh* variant was named *amhy*, and the *amh* shared by both sexes was named *amhx*. The association analysis using both females and males of the genotypic sex inferred by the presence/absence of *amhy* found complete association with phenotypic sex and genotype. Gene expression analysis in larvae derived from a single-pair progeny by quantitative real-time PCR detected *amhy* transcripts in the larval trunks between 20 and 100 days after hatching only in XY larvae. Localization of *amhy* by *in situ* hybridization was detected in presumptive Sertoli cells of XY gonads. Expression of *amhx* was almost undetectable in both XX and XY genotypes. Loss of Amh function by CRISPR-Cas9 induced male-to-female sex reversal, indicating that this gene was necessary for the masculinization of XY individuals. In conclusion, the complete linkage of *amhy* with males, its early expression in XY gonads before testicular differentiation, and the induction of

sex reversal by loss-of-function mutation support the view that *amhy* is the sex-determining gene in this species.

KEYWORDS

**testis-determining gene, *amhy*, *amhx*, sex determination, müllerian-inhibiting substance**

## Introduction

Sex differentiation of the gonads in gonochoristic species is a binary, antagonistic process determined by the balance between female-promoting and male-promoting factors, which culminates in the commitment of the undifferentiated gonad into either an ovary or testis (Hattori et al., 2020). In fish, this process is influenced by genotypic and environmental factors wherein the strength of one over another may differ greatly according to the species or taxonomic group (Fernandino et al., 2013).

Research on genotypic factors of sex determination in teleosts has unraveled a large repertoire of sex-determining genes, including the canonical transcription factors with DM- or SOX-binding domains as well as some unusual players such as the immune-related *sdY* of salmonids (Yano et al., 2013) and the steroidogenic enzyme *hsd17b1* in *Seriola quinqueradiata* (Koyama et al., 2019). Despite such variation, there is seemingly a high likelihood that members of the TGF-beta gene superfamily will be recruited as sex-determining genes, particularly those related to Amh-signaling. Y-specific *amh* duplication (namely *amhy*) has been reported in some Atheriniformes (Hattori et al., 2012; Yamamoto et al., 2014; Bej et al., 2017; Hattori et al., 2019), in Esociformes (Pan et al., 2019), and in the Cichliformes Nile tilapia (Li et al., 2015). In some fugu species (*Takifugu* sp.), *amhrII*, the receptor that is supposed to bind to *amh*, has been proposed as the key sex-determining gene (Ieda et al., 2018).

Despite the presence of well-known genotypic sex determinants of sex, in some species environmental cues can overcome the course of sex differentiation, inducing the appearance of male-to-female or female-to-male sex-reversed fish (Hattori et al., 2020). Currently, both (genotypic and environmental) sex determination systems are considered as extremes of a continuum rather than two independent and mutually exclusive mechanisms (Yamamoto et al., 2014) because they seem to coexist even in wild populations (Baroiller and D'Cotta, 2016; Miyoshi et al., 2020). The physiological and molecular mechanisms governing thermolabile sex determination have advanced remarkably since the discovery of the link between the stress hormone cortisol and masculinization in teleosts (Hattori et al., 2009; Hayashi et al., 2010; Yamaguchi et al., 2010). Cortisol is produced by the adrenal glands and its release is stimulated by Crh signaling (the ligand Crhb and the receptors Crhr1 and Crhr2) during masculinization of XX genotypes in Japanese medaka (Castañeda Cortés et al., 2019), suggesting that the central nervous system has an important role in gonadal sex determination induced by the environment.

Japanese flounder (*Paralichthys olivaceus*) is a marine flatfish with high commercial importance for fisheries and aquaculture in Northeast Asia (Japan, Korea, and China). Sex determination is controlled both genotypically by a XX-XY male heterogametic system (Yamamoto, 1999) and environmentally by sex steroids and water temperature (Hayashi et al., 2010). The latter can affect sex determination by driving undifferentiated gonad into either ovary or testis when exposure occurs before fish reach 40 mm of total body length (Yamamoto, 1999). According to this study, temperatures below 17.5°C or above 22.5°C were associated to high percentage of males. All XX progeny production by gynogenesis or crosses between XX genotypes (neomales with normal females) has been adopted to improve female production because females generally have better growth rates than males after sexual maturation; other female-specific traits such as lower susceptibility to diseases are also factors behind the preference for females in most fish species (Peterson and David, 2012). However, despite the absence of male genetic factors in these progeny, some individuals undergo natural female-to-male sex reversal, yielding undesired phenotypic males among the XX progenies (Hayashi et al., 2010). Warm temperature of approximately 22°C is likely the major factor that affects masculinization, but other stressors such as background color may also be involved, as was demonstrated in *Paralichthys lethostigma* (Mankiewicz et al., 2013).

Thus, basic information on genotypic sex determination is important for better understanding the genetic molecular mechanisms governing sex determination and for providing molecular tools for the development of reproductive biotechnologies in *P. olivaceus*. Considering that environmental sex determination is also strong in this species and that sex ratio distortions were found to occur even in wild populations of *P. lethostigma* (Honeycutt et al., 2019), the availability of a sex-determining gene could be instrumental for research on the impacts of global warming/climate change on natural populations of Japanese flounder. We designed this study with the ultimate goal of unravelling the genotypic sex determination mechanism in *P. olivaceus*.

For this, information on sex-linked microsatellite markers and genome data were combined and used to identify a candidate sex-determining gene in *P. olivaceus*. The association between the candidate gene and phenotypic sex was investigated in farmed and wild individuals and its mRNA expression pattern during sex differentiation was analyzed.

# Materials and methods

## Source of fish used in this study

To identify sex-linked microsatellite markers, we generated a family (family-A) using a gynogenetic female (A-female: XX), produced by retention of the second polar body (meiotic gynogenesis), and a normal male (A-male: XY), derived by crosses between normal diploid fish at the Tottori Prefectural Fisheries Experimental Station (Tottori Prefecture, Japan). Larvae were treated with a low dose (0.3 μg/g of food weight) of estrogen (estradiol-17$\beta$; $E_2$) from 41 to 70 dah (days after hatching) when they were an average total length of 20–50 mm. The purpose of the estrogen treatment was to inhibit spontaneous female-to-male sex reversal, as was described previously (Yamamoto, 1995; Yamamoto, 1999). The larvae were reared at a constant water temperature of 20°C. Nine months after hatching, the sex of the progeny was determined by visual inspection of the gonads. The proportion of phenotypic females and males was 46.2% ($n$ = 36) and 53.8% ($n$ = 42), respectively. A genetic map was constructed by genotyping the parents and their F1 offspring.

To fine map the sex-determining locus, we generated another family (family-B) by crossing a sex-reversed female (B-female: XY) and a sex-reversed male (B-male: XX), both of which were produced by hormone treatment and screened using the microsatellite markers *Poli30MHFS* and *Poli31MHFS*. The sex of progeny was determined by visual inspection of the gonads at 8 months of age. A total of 38 family-B individuals (25 males and 13 females) was used for the mapping. We used sex-reversed individuals because of the different rates of recombination between females and males, as described previously (Sakamoto et al., 2000; Castaño-Sánchez et al., 2010); females have higher rates of recombination near centromeres, whereas males have higher rates of recombination near telomeres. We expected that a linkage analysis using family-A and family-B families would allow a more accurate characterization of the sex-determining locus in *P. olivaceus*.

For the association analysis between the presence of the Y-specific *amh* gene (hereafter *amhy*) and phenotypic sex, we used wild-caught juveniles from the estuarine area of the Yurugawa River (Maizuru Bay, Japan). The juveniles were more than 50 mm (standard length) and supposedly had their phenotypic sex already determined in the wild. They were collected by seine net and reared for one more year at the National Research Institute of Aquaculture, Fisheries Research Agency (Oita, Japan), where the gonadal sex of all the individuals was determined by visual inspection under a stereomicroscope.

For the genetic analysis, genomic DNA was extracted from the caudal fin of each individual using either phenol/chloroform or a DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany).

## Rearing conditions of experimental fish

Samples for the gene expression analysis were obtained from a single pair cross, which was selected due to the presence of a single-nucleotide polymorphism (SNP) in exon 2 of the candidate sex-determining gene (see description in the sequencing section). Egg incubation and larvae rearing were conducted in UV-treated sea water at 18°C for up to 100 dah under a constant light cycle (16-h light/8-h dark). Larvae/juveniles were measured and sampled every five or 10 days between 20 and 80 dah (Supplementary Table S1). The remaining fish were sampled and their phenotypic sex was determined by gonadal histology (100 dah). Caudal fins were fixed in ethanol for sex genotyping and trunks were fixed either in RNALater™ (Ambion Inc., Austin, United States) or 4% paraformaldehyde for expression analysis.

## Genotyping by microsatellite markers and linkage analysis of sex-determining locus

Microsatellite genotyping was performed in a 10-μl reaction volume containing 0.2 pmol/μl of unlabeled primer and 0.03 pmol/μl of end-labeled with [γ-$^{33}$P]ATP using T4 polynucleotide kinase, plus 1× buffer, 0.2 mM dNTP, 1% bovine serum albumin, 0.02 U of *Taq* DNA polymerase, and 50 ng template DNA. A specific annealing temperature was used for each microsatellite marker. The PCRs were run on a GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, United States) under the following conditions: initial denaturation at 95°C for 2 min, followed by 35 cycles of 30 s at 95°C, 1 min at the annealing temperature, 1 min at 72°C, and final extension of 3 min at 72°C. The amplification products were mixed with 10 μl of loading buffer (95% formamide, 10 mM EDTA, 0.05% bromophenol blue, and xylene cyanol), denatured for 10 min at 95°C, and quickly cooled on ice. Then, 2 μl of each sample was loaded onto a 6% denaturing polyacrylamide gel (19:1 ratio acrylamide: bisacrylamide). After electrophoresis, the gels were dried on a standard gel drier for 30 min and exposed to imaging plates (FUJIFILM, Tokyo, Japan) overnight. The imaging plates were scanned with a Bio-image Analyzer, BAS1000 (FUJIFILM).

The linkage analysis to identify the sex-determining locus of Japanese flounder was carried out in three steps. In step 1, we analyzed the segregation of paternally inherited alleles of 63 microsatellite markers in 44 individuals (22 males and 22 females) of family-A to determine whether they were sex

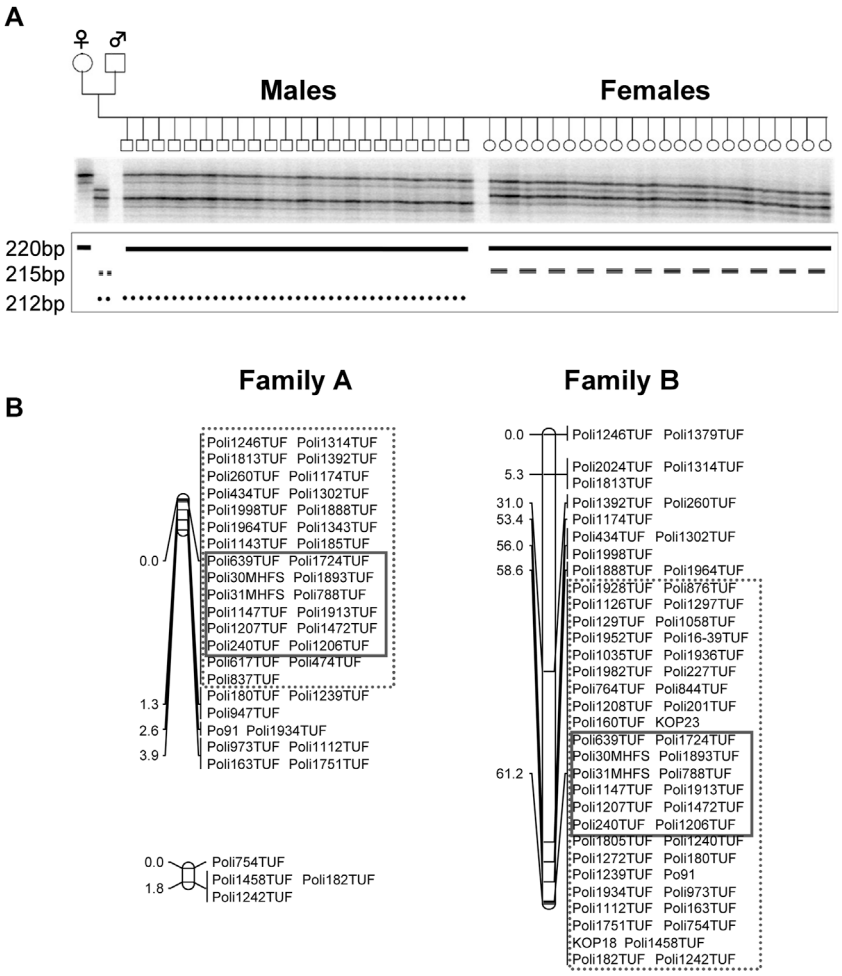



FIGURE 1
Autoradiograph of Poli185TUF in JF9 in the A-family at step 1 **(A)**. The lower 212 bp allele from the male parent was the unique allele inherited by the males of this family. JF9-male linkage map in Family A and JF9 linkage map in Family B **(B)** Map distances between markers are shown in centimorgans (cM). 35 markers in the enclosure with a dotted line were mapped close to sex-determining locus in A-family, while 46 markers in the enclosure with a dotted line were mapped in family **(B)**. Common 12 markers in the enclosure with a straight line were detected as closed markers to the sex-determining locus.

linked. These markers were selected to represent all linkage groups (LGs) of the genetic linkage map of Japanese flounder (Castaño-Sánchez et al., 2010). We analyzed the segregation of paternally inherited alleles of all the markers to determine whether they were sex linked. The linkages between the genotypes at each locus and phenotypic sex were tested using the Map Manager QT software program (Manly and Olson, 1999). Markers with LOD scores above 3.0 were considered potential sex-linked markers. In step 2, we used the potential sex-linked markers together with other markers that were mapped to the same LG to genotype all the individuals of family-A (42 males and 36 females). In step 3, for the fine mapping of the sex-determining locus, we used 59 markers on the same LG (LG 9) to genotype individuals of family-B (25 males and 13 females). Maps were drawn using MapChart v2.0 (Voorrips, 2002).

## Sequence analysis of scaffolds containing sex-linked microsatellite markers

We constructed the draft genome of *P. olivaceus* captured from the wild population of Wakasa Bay (Kyoto, Japan) by whole-genome shotgun assembly using Illumina short reads and PacBio long reads. The genome assembly has been deposited in the DDBJ database under accession numbers BRVK01000001–BRVK01002790.

The sex-linked microsatellite markers were searched against our draft scaffolds of the *P. olivaceus* genome using BLASTN (Altschul

TABLE 1 Segregation of paternally inherited alleles of 63 microsatellite markers in 44 individuals (22 males and 22 females) of family A. The marker in bold represent the sex-determining locus.

| Linkage group | Locus | *P* | LOD score | Linkage group | Locus | *P* | LOD score |
|---|---|---|---|---|---|---|---|
| 1 | Poli110TUF | 0.32 | 0.22 | 11 | Poli174TUF | 0.01 | 1.41 |
| | Poli41MHFS | 0.32 | 0.22 | | Poli176TUF | 0.12 | 0.52 |
| | Poli9-67TUF | 0.13 | 0.50 | | Poli132TUF | 0.06 | 0.74 |
| 2 | Poli23TUF | 0.54 | 0.09 | 12 | Poli38MHFS | 0.76 | 0.02 |
| 3 | Poli13TUF | 0.07 | 0.74 | | Poli179TUF | 0.35 | 0.20 |
| | Poli188TUF | 0.76 | 0.02 | | Poli149TUF | 0.36 | 0.17 |
| | Poli24MHFS | 0.19 | 0.37 | 13 | Poli145TUF | 0.22 | 0.33 |
| | Poli13MHFS | 0.54 | 0.09 | | Poli17MHFS | 0.22 | 0.33 |
| | Poli18-51TUF | 0.07 | 0.74 | | Poli204TUF | 0.54 | 0.09 |
| 4 | Poli148TUF | 0.22 | 0.33 | 14 | Poli-RC47-TUF | 0.53 | 0.09 |
| | Poli128TUF | 0.54 | 0.09 | | Poli120TUF | 0.22 | 0.33 |
| | Poli19TUF | 0.75 | 0.02 | 15 | Poli9-8TUF | 1.00 | 0.00 |
| | Poli14MHFS | 0.76 | 0.02 | | Poli121TUF | 0.76 | 0.02 |
| | Poli142TUF | 0.75 | 0.02 | | Poli-RC15-35-TUF | 0.22 | 0.33 |
| 5 | Poli43TUF | 0.76 | 0.02 | 16 | Poli114TUF | 0.22 | 0.33 |
| | Poli169TUF | 0.75 | 0.02 | | Poli199TUF | 0.76 | 0.02 |
| | Poli151TUF | 0.55 | 0.09 | 17 | Poli127TUF | 0.75 | 0.02 |
| 6 | Poli190TUF | 0.21 | 0.35 | | Poli11TUF | 0.76 | 0.02 |
| | Poli143TUF | 0.36 | 0.17 | 18 | Poli147TUF | 0.36 | 0.17 |
| | Poli107TUF | 0.22 | 0.33 | | Poli16-79TUF | 1.00 | 0.00 |
| 7 | Poli18-55TUF | 0.36 | 0.17 | 19 | Poli108TUF | 1.00 | 0.00 |
| | Poli113MHFS | 1.00 | 0.00 | | Poli186TUF | 0.22 | 0.33 |
| | Poli50TUF | 1.00 | 0.00 | 20 | Poli123TUF | 0.76 | 0.02 |
| 8 | Poli202TUF | 0.15 | 0.43 | | Poli9-58TUF | 0.00 | 0.00 |
| | Poli126TUF | 0.36 | 0.17 | 21 | Poli7 MHFS | 0.76 | 0.02 |
| 9 | Poli182TUF | 0.54 | 0.09 | | Poli109 MHFS | 0.22 | 0.33 |
| | **Poli185TUF** | **0.00** | **13.25** | 22 | Poli2TUF | 0.06 | 0.78 |
| | Poli180TUF | 0.00 | 11.17 | 23 | Poli150TUF | 0.55 | 0.09 |
| 10 | Poli144TUF | 0.55 | 0.09 | | Poli18-42TUF | 0.36 | 0.17 |
| | Poli13-2TUF | 0.55 | 0.09 | | Poli56TUF | 0.09 | 0.61 |
| | Poli37MHFS | 0.76 | 0.02 | | Poli-RC27-TUF | 0.02 | 1.15 |
| | | | | 24 | Poli183TUF | 0.36 | 0.17 |

et al., 1990). Then, the sequences of the aligned scaffolds were analyzed using GENSCAN (Burge and Karlin, 1997) to predict the genes. The potential roles of the identified genes were inferred based on the UniProt database annotations (Bateman et al., 2021).

## Genome resequencing and screening for sex-linked SNPs

The genomes of 17 wild-caught *P. olivaceus* individuals were sequenced on an Illumina HiSeq X platform with 150-bp paired-end reads. The library construction and sequencing were conducted by BGI (Shenzhen, China). Low quality sequences were trimmed off using Trimmomatic v3.6 (Bolger et al., 2014) and mapped against the *amh*-containing scaffolds (786.920 bp) as reference using BWA-mem v0.7.12 (Li and Durbin, 2009) with default parameters. SNP detection was performed using Samtools mpileup (Garrison and Marth, 2012) with the following parameters: minimum coverage = 10, maximum coverage = 50. After removing mutations derived from low quality sequences using Vcffilter of Vcflib software, the high precision polymorphisms were filtered using VCFtools with parameters mac7, max0mac7, and max-missing-count 0 (Danecek et al., 2011).

To detect SNPs in the *amh* gene, wild-caught *P. olivaceus* (20 females and 8 males) were analyzed by direct sequencing. PCR fragments were amplified using specific primers and sequenced on an ABI PRISM 3100 capillary sequencer

(Applied Biosystems, Foster City, United States) using the BigDye Terminator method. Sequences were analyzed using GENETYX v12.0 (GENETYX, Tokyo, Japan).

## Comparative sequence analysis of amh in XX and XY *P. olivaceus*

To amplify the *amh* gene sequences, including exons, introns, and the 5′- and 3′-flanking regions, DNA was extracted from the caudal fin of XX and XY fish using a Qiagen DNA Extraction Kit (Qiagen, Hilden, Germany). PCRs were performed in a total volume of 10 µl with Ex Taq polymerase (Takara Bio, Shiga, Japan) and the primers described in Supplementary Table S2. Amplification was performed in a Thermocycler T professional BASIC 96 Gradient (Biometra GmbH, Goettingen, Germany). Fragments were excised from the gel, purified, and cloned into a pGEM-T easy vector (Promega, Corporation, Madison, WI). Sequencing was performed as described in the previous section.

## Amplification of 5′ and 3′ untranslated regions by RACE-PCR

To isolate the 5′ and 3′ untranslated regions of the Y-specific *amhy* gene, gonads of a genetic male larva (55 dah) expressing *amh* (previously assessed) were used. To isolate the 5′ and 3′ untranslated regions of the X-specific *amh* gene (*amhx*), ovary tissues from an adult XX female were used. The extracted RNA was reverse transcribed using a Clontech SMART rapid amplification of cDNA ends (RACE) Amplification Kit (BD Biosciences, Franklin Lakes, United States) and used for PCRs with gene-specific primers (Supplementary Table S3) following the manufacturer's instructions. The PCR-amplified fragments were cloned and sequenced, as described in the previous section.

## Sex genotyping by 5′ upstream region amplification of the *amh* gene

Caudal fin samples were treated with 50 mM NaOH for 5 min at 95°C, equilibrated with Tris-HCl (pH 8.0), and briefly centrifuged to isolate the DNA template. The aqueous phase containing DNA was used for genotyping using KOD Fx Neo (TOYOBO, Osaka, Japan) and primers amhFw1 5′-AAGTTCAGTTCAGTTGCACAGC-3′ and amhRv1 5′-CTTGACAAACAGGGCATGAATA-3′ under the following conditions: initial denaturation at 94°C for 2 min, followed by 35 cycles of 30 s at 95°C, 30 s at 64°C, and 1 min at 68°C, and final elongation of 3 min at 68°C. The amplified products were electrophoresed in 1% agarose gel and visualized by ethidium bromide staining. Samples with two bands (931 bp and 687 bp) were scored as genotypic males (XY) and those with a single band (931 bp) were scored as genotypic females (XX).

## Expression analysis by qRT-PCR during sex differentiation

Total RNA was extracted from larval trunks using an RNeasy mini kit (QIAGEN K.K., Tokyo, Japan) following the manufacturer's instructions. RNA samples (1 µg) were treated with deoxyribonuclease I amplification grade (Invitrogen) and reverse transcribed using SuperScript III RNase H-Reverse Transcriptase (Invitrogen) with oligo(dT) 12-18 following the manufacturer's instructions. The quantitative real-time PCRs (qRT-PCRs) were performed in 20-µl reaction volumes using either Premix Ex Taq™ and specific TaqMan minor-groove binder (MGB) probes for both *amhy* and *amhx* or TB Green® Premix Ex Taq™ for *cyp19a1a* and *amhrII* (both from Takara Bio Inch, Shiga, Japan). Amplification was done using 1 µl of first strand cDNA (approximately 25 ng) and 5 pmol of each primer (Supplementary Table S3) in a StepOne Plus Real-Time PCR system (Applied Biosystems, Foster City, United States). Transcript abundance was quantified using the standard curve method with four dilution points and normalized against *elf1a* values.

## Histological determination of sex ratios and localization of *amh* transcripts by *in situ* hybridization

Larval trunks were collected at the end of the experimental period and fixed in 4% paraformaldehyde solution overnight, dehydrated in an ascending ethanol series, and stored in absolute ethanol. Samples were embedded in Paraplast Plus (McCormick, St. Louis, United States), sectioned transversally at a thickness of 6 µm, and mounted on glass slides. Sections were stained with hematoxylin-eosin and observed under a microscope for determination of gonadal sex.

*In situ* hybridization was performed in larvae collected before (35 dah) and after (100 dah) the onset of histological differentiation of the gonads using the procedures described previously (Sarida et al., 2019). The probe was synthesized using the same primers as those used in a previous study (Yoshinaga et al., 2004). Hybridization was conducted in the transverse sections overnight at 63°C. NBT/BCIP was used for signal detection in accordance with the manufacturer's recommendations (Roche Diagnostics, Basel, Schweiz).

10.3389/fgene.2022.1007548

TABLE 2 Detailed list of scaffolds blasted against 12 sex-linked microsatellite markers.

| Marker name | Scaffold # | Identity | Start | End | E-value | Accession number |
|---|---|---|---|---|---|---|
| Poli1206TUF | 318 | 97.235 | 193,315 | 193,099 | 5.7E-99 | BRVK01000318 |
| Poli240TUF | 717 | 97.561 | 80,385 | 80,917 | 0 | BRVK01000717 |
| Poli1472TUF | 801 | 93.269 | 105,926 | 106,237 | 3.7E-122 | BRVK01000801 |
| Poli1207TUF | 126 | 94.86 | 50,689 | 50,902 | 7.24E-88 | BRVK01000126 |
| Poli1913TUF | 346 | 100 | 672,910 | 672,650 | 5.3E-135 | BRVK01000346 |
| Poli1147TUF | 362 | 98.319 | 113,310 | 113,547 | 3.8E-116 | BRVK01000362 |
| Poli788TUF | 410 | 97.674 | 297,718 | 297,289 | 0 | BRVK01000410 |
| Poli31MHFS | 752 | 98.374 | 237,106 | 237,228 | 1.59E-54 | BRVK01000752 |
| Poli1893TUF | 274 | 98.106 | 76,773 | 76,514 | 3.3E-127 | BRVK01000274 |
| Poli30MHFS | no hits | | | | | - |
| Poli1724TUF | 301 | 92.808 | 82,166 | 81,890 | 4.7E-111 | BRVK01000301 |
| Poli639TUF | 970 | 94.407 | 174,735 | 174,147 | 0 | BRVK01000970 |

TABLE 3 Localization of 15 SNPs that showed association with sex in relation with the predicted genes from the 786.920 bp scaffold. Gene #101 corresponds to the *amh* gene.

| SNP type | | Gene # | Region | Type of substitution | Position in the scaffold |
|---|---|---|---|---|---|
| XX | XY | | | | |
| T/T | T/A | 83 | Intron | - | 47,837 |
| C/C | C/G | 96–97 | Intron | - | 228,359 |
| T/T | T/A | | Intron | - | 229.620 |
| T/T | T/A | 98 | Intron | - | 234,671 |
| C/C | C/T | 100 | Intron | - | 242,211 |
| C/C | C/T | | Intron | - | 242,984 |
| A/A | A/T | 101 | Intron | - | 243,573 |
| G/G | G/A | | Exon | Synonymous | 244,664 |
| C/C | C/T | | Exon | Synonymous | 245,185 |
| T/T | T/A | | Intron | - | 245,289 |
| A/A | A/T | | Intron | - | 247,694 |
| T/T | T/C | 102 | Intron | - | 255,206 |
| G/G | G/A | | Exon | Synonymous | 257,160 |
| T/T | T/A | | Intron | - | 258,300 |
| G/G | G/A | | Intron | - | 260,102 |

## CRISPR-Cas9-mediated loss-of-function of the *amh* gene

Synthetic CRISPR RNAs (crRNAs) and trans-activating crRNA (tracrRNA) were obtained from Fasmac Co., Ltd. (Kanagawa, Japan). The sequences of the two crRNAs for *amh* were Target 1 5′-CUAUCUGCAGCUCGUAGGUAguuuuagagcuaugcuguuuug-3′ and Target 2 5′-AAUUAAAAAGCACAUUUUGaguuuuagagcuaugcuguuuug-3′. These sequences were designed on the basis of Sawamura et al. (Sawamura et al., 2017). Two crRNAs (250 ng/μl), tracrRNA (500 ng/μL), and Cas9 protein (750 ng/μl; Toyama, Nippongene) were mixed and immediately injected into fertilized eggs using a microinjector (IM-9B, Narishige, Tokyo, Japan). Two crRNAs and Cas9 protein without tracrRNA were injected as experimental controls.

The artificial fertilization, microinjection, and rearing of the experimental fishes were performed at the Nansei Field Station, Japan Fisheries Research and Education Agency. Eggs and sperm from three-year-old females and males were used. They were mixed in

a plastic beaker sampler and activated with sterilized seawater (18°C) for artificial fertilization. At 200 dah, the gonads were isolated from each fish; one lobule was processed for histological analysis and the other was used for sex genotyping. All the fish were sampled after ensuring that they had been completely euthanized by an overdose of 2-phenoxyethanol (Wako Chemicals).

## Sequence analysis of mutations

Genomic DNA was extracted from the isolated gonads using ISOGEN® (Nippongene) according to the manufacturer's instructions. Genomic PCRs were carried out using AmpliTaq Gold® (Applied Biosystems) with the following *amh*-specific primers: Forward 5′-TTTCTTCTC CTGAAGGCCC-3′ and Reverse 5′-ATTAGCTGTCACAGC AGCAG-3′. The PCR conditions were as follows: pre-heating at 95°C for 2 min, followed by 35 cycles at 95°C for 15 s, 59°C for 30 s, 72°C for 2 min, and final extension at 72°C for 5 min. The amplified PCR fragments were subcloned using a TA PCR Cloning Kit (BioDynamics Laboratory Inc., Tokyo, Japan), then sequenced as described previously.

## Statistical analysis

The significance of differences between groups was determined by one-way ANOVA followed by the Tukey test for gene expression using GraphPad Prism v6.0 (GraphPad Software, San Diego, United States). Differences were considered as statistically significant at $p < 0.05$.

# Results

## Linkage analysis and fine mapping of sex-determining locus

The sex-determining locus was identified in LG 9 (Figure 1A and Table 1) in family-A. We used the sex-linked markers identified in step 1 of the linkage analysis and additional microsatellite markers in LG 9 in all individuals of family-A and found 29 markers that were perfectly linked with the sex-determining locus with no recombination in family-A. In the fine mapping using family-B, we identified 46 markers that were tightly linked with the sex-determining locus with no recombination. Among them, we

TABLE 4 Detailed characterization of SNPs detected in the coding regions of *amh* gene in XX and XY genotypes.

| Region | | Exon 2 | | Exon 3 | Exon 4 | Exon 6 | Exon 7 | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | CDS | | 3′ UTR | |
| SNP ID | | #01 | #02 | #03 | #04 | #05 | #06 | #07 | #08 | #09 |
| Position from start codon (bp) | | 289 | 493 | 631 | 924 | 1445 | 2014 | 2034 | 2544 | 2546 |
| Females (XX) | F01 | C/C | T/T | C/T | C/C | G/G | C/C | C/C | G/G | A/A |
| | F02 | C/C | T/T | C/T | C/T | G/G | C/C | C/C | G/G | A/A |
| | F03 | C/C | T/T | C/T | DEL | G/A | C/C | C/C | G/G | A/A |
| | F04 | C/C | T/T | C/C | C/C | G/G | C/C | C/C | G/G | A/A |
| | F05 | C/C | T/T | C/C | C/C | G/G | C/C | C/C | G/G | A/A |
| | F06 | C/C | T/T | C/C | C/C | G/G | C/C | C/C | G/G | A/A |
| | F07 | C/C | T/T | C/C | C/C | G/G | C/C | C/C | G/G | A/A |
| | F08 | C/C | T/T | C/C | C/C | G/G | C/C | C/C | G/G | A/A |
| | F09 | C/C | T/T | C/C | C/C | G/G | C/C | C/C | G/G | A/A |
| Males (XY) | M01 | C/T | T/G | C/T | C/T | G/A | C/T | C/T | G/T | A/T |
| | M02 | C/T | T/G | C/T | C/T | G/A | C/T | C/T | G/T | A/T |
| | M03 | C/T | T/G | C/T | C/T | G/A | C/T | C/T | G/T | A/T |
| | M04 | C/T | T/G | C/T | C/T | G/A | C/T | C/T | G/T | A/T |
| | M05 | C/T | T/G | C/T | C/T | G/A | C/T | C/T | G/T | A/T |
| | M06 | C/T | T/G | C/T | C/T | G/A | C/T | C/T | G/T | A/T |
| | M07 | C/T | T/G | C/T | C/T | G/A | C/T | C/T | G/T | A/T |
| | M08 | C/T | T/G | C/T | C/T | G/A | C/T | C/T | G/T | A/T |

detected 12 markers that were perfectly linked with the sex-determining locus and were commonly shared by both families (Figure 1B and Table 2).

## Sequence analysis of scaffolds containing sex-linked microsatellite markers

The *P. olivaceus* genome assembly yielded 2,790 scaffolds longer than 1,000 bp. The longest scaffold was 1.81 Mb with N50 of 357.0 kb, the mean scaffold length was 217.0 kb, and the total size of the assembly was 605.6 Mb. The 12 sex-linked microsatellite markers were distributed in 11 of the scaffolds (Table 2). A total of 181 genes were predicted in these scaffolds based on the UniProt database annotations (Supplementary Table S4). Among them, we selected the *amh* gene because of its known involvement with male sex determination in some teleosts.

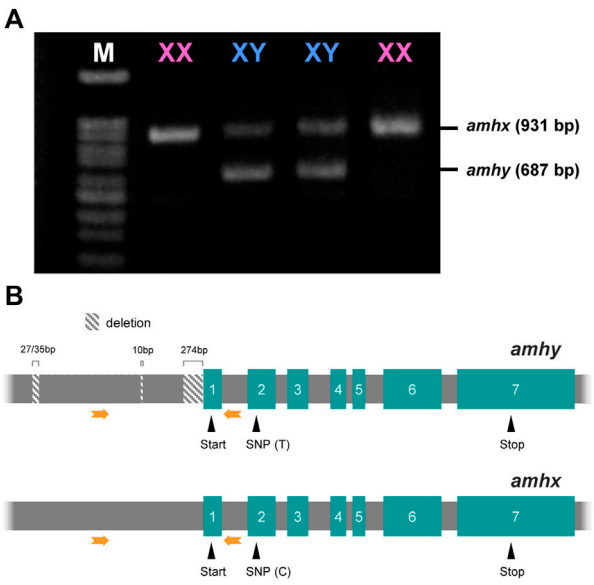



FIGURE 2
Amplification pattern of 5′ upstream region and the structure of *amhy* and *amhx* genes in *P. olivaceus*. **(A)** Agarose gel electrophoresis of *amhx* and *amhy* 5′ upstream region amplified by PCR in XX and XY individuals, phenotypic females and males, respectively; *amhy* band is shorter than *amhx* due to deletions in the presumptive promoter. **(B)** Structures of *amhy* and *amhx* genes showing the position of three *amhy*-specific deletions in the 5′ upstream region, both start and stop codons, and a SNP in the second exon; thymine (T) and cytosine **(C)** in exon 2 were highly linked to *amhy* and *amhx*, respectively. Green boxes represent the exons with the respective number and grey boxes in between represent the introns. Orange arrows indicate the position of primers used in **(A)**

## Linkage analysis between SNPs and the sex phenotype

We detected 69 SNPs in the *amh* gene and 15 of them were highly associated with maleness (Table 3). The in-depth analysis of coding and non-coding regions of the *amh* gene in XX and XY individuals detected nine SNP markers in exons 2, 3, 4, 6, and 7, but none of them were synonymous substitutions (Table 4).

## Size polymorphism in the promoter regions of the *amh* gene

Amplification of the region 5′ upstream of the start codon and gel electrophoresis of the products showed that there was a sex-specific polymorphism in the banding pattern; genetic females had one band and genetic males had two bands. One of the two bands in the males was the same size as the band in the females; the other band was slightly shorter (Figure 2A). The sequence analysis showed that the male-specific band had deletions in three segments, one of 274 bp, one of 10 bp, and one of 27 bp or 32 bp at positions −14 bp, −449 bp, and −972 bp, respectively (the positions correspond to the *amhx* gene sequence, which was used as the reference; Figure 2B). The sequences of the coding regions of the *amhy* and *amhx* genes were almost identical, indicating the presence of two alleles for *amh* that differ in the presumptive promoter and that theoretically encode identical proteins because the SNPs detected in some of the exons were synonymous substitutions. No other *amh* homologs were identified in other chromosomes based on searches against the *P. olivaceus* genome database provided in this study.

## Association analysis between *amhy* and the male phenotype

The association analysis using females and males for which the genotypic sex was inferred by the presence/absence of *amhy* showed complete association with phenotypic sex and genotype in captive-reared individuals used to identify the microsatellite markers. Complete association was also found in a wild-caught population from Maizuru Bay, the Sea of Japan; 20 females where negative for *amhy* and the eight males were *amhy*-positive. Considering that *P. olivaceus* has an XX-XY sex determining system, it can be surmised that the first allele is on the Y chromosome and its counterpart is on the X chromosome. For this reason, the lighter band was named *amhy* (Y chromosome-linked *amh*), whereas the first, upper band was named *amhx* (X chromosome-linked *amh*).

The synonymous SNP that was found in the 39$^{th}$ bp of exon 2, showed a high but not full association with two of the *amhy*-specific deletions; 94.74% of fish had the nucleotide T in *amhy* and C in *amhx*.

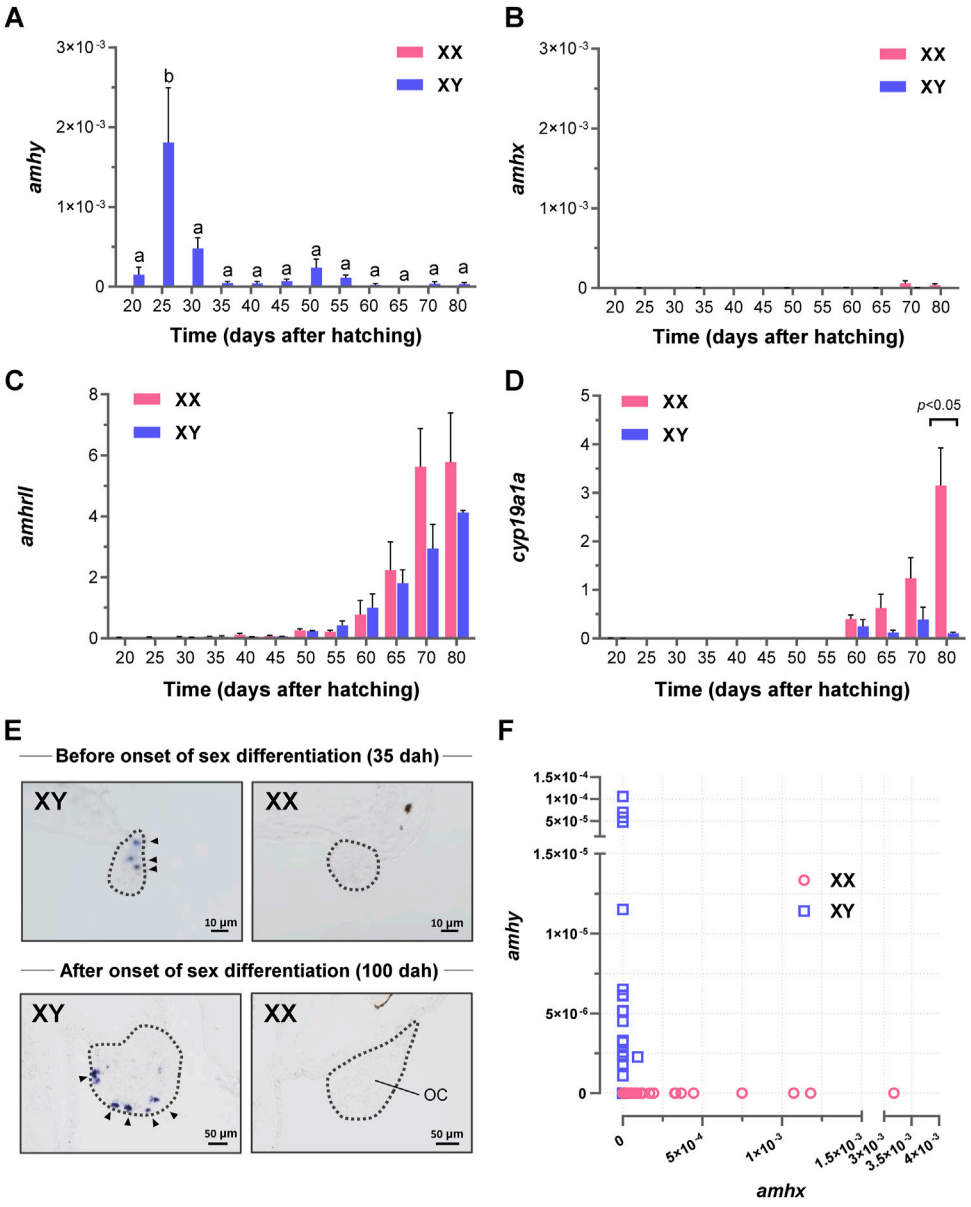



FIGURE 3
Expression analysis of *amhy, amhx, amhrII*, and *cyp19a1a* mRNA during sex differentiation in *P. olivaceus*. Expression profiling by qRT-PCR in larvae trunks showed **(A)** *amhy* expression only in XY genotypes with a peak at 25 dah. **(B)** *amhx* was almost undetectable during the same period; weak expression was detected at 70 and 80 dah. **(C,D)** Quantitative analysis of *amhrII* and *cyp19a1a* mRNAs during the same period; sexually dimorphic expression was detected only for *cyp19a1a*, at 80 dah XX samples. **(E)** Spatial localization of *amh* riboprobe restricted to XY gonads before and after the onset of sex differentiation; at 100 dah, signals were recognized at the ventral edge of the XY gonad, where germ cells are distributed to form cysts. In XX gonads, which had a clear ovarian cavity (OC), no detectable signals of *amh* were found. **(F)** the relationship between *amhy* and *amhx* expression in XY in relation to *amhx* in XX; although absolute expression values were generally lower than *amhy*, for some reason more XX individuals were expressing *amhx* than XY. Differences were considered as significant for $p < 0.05$ by One-Way ANOVA with Turkey post-test.

## Expression analysis of *amhy* during sex differentiation

Expression analysis by qRT-PCR detected *amhy* transcripts from 20 to 80 dah in the XY samples with a peak at 25 dah (Figure 3A), whereas *amhx* mRNA expression was very low or undetectable during the same period in both genotypes (Figure 3B). The expression of *amhrII* was detected from 25 dah and increased gradually from 55 dah onwards (Figure 3C), and that of *cyp19a1a* was

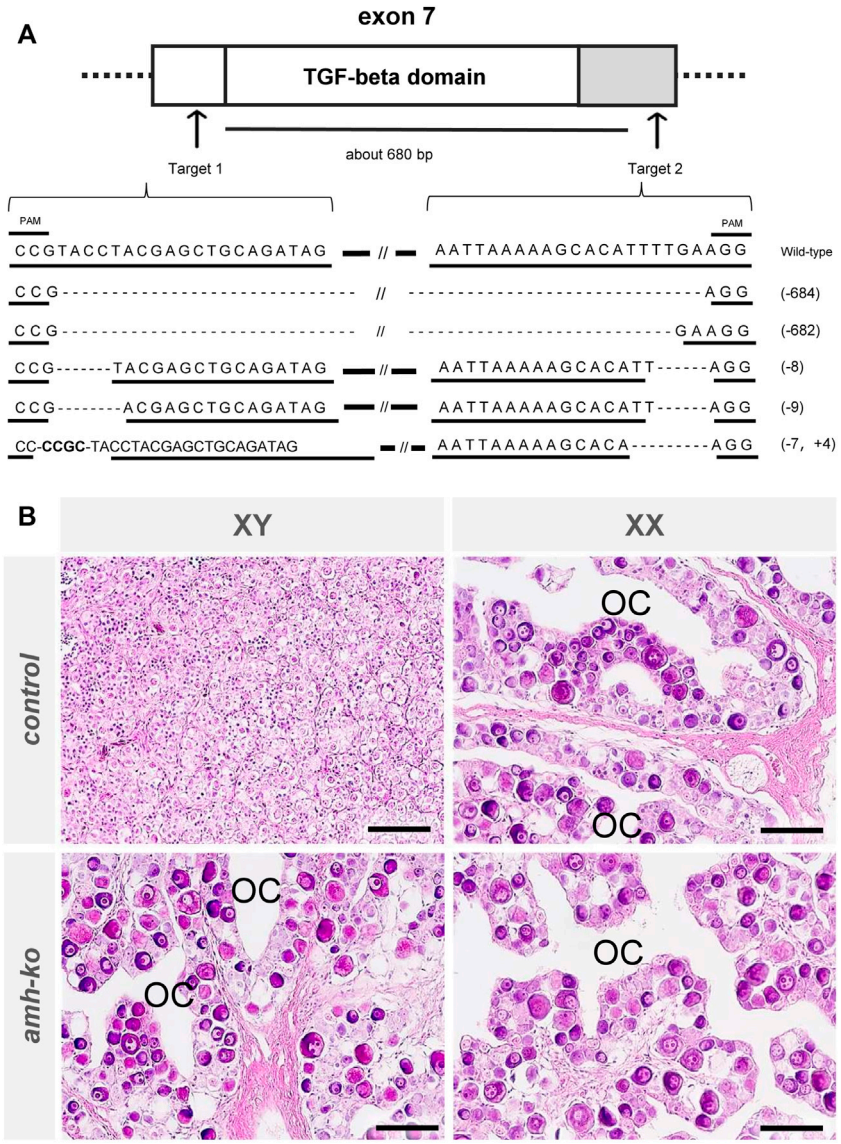



FIGURE 4
Production of *amh* mutant Japanese flounder using CRISPR/Cas9 system. **(A)** Schematic positions of crRNAs targeted on *amh* locus, showing exons (white boxes), untranslated regions (gray boxes), introns (horizontal lines) and target sequences of the crRNAs with CRISPR/Cas9 system (arrows). The numbers between target sequences indicate approximate deletion size by CRISPR/Cas9 system. Sequences of wild-type *amh* and typical mutation of *amh* mutant with the rate in two target regions. The underlined sequences indicate the two crRNA target sequences. The horizontal lines and dashed line indicate wild-type sequences and deletions, respectively. Slashes indicate omission of long sequences. The number in parentheses indicates the size of deletion (-) or insertion (+). **(B)** Histological images of gonads in wild-type and *amh* mutant Japanese flounder at 200 days after hatching (dah). In controls, genetic male gonad showed typical testis containing many spermatogonia, and genetic female gonad showed typical ovary harboring many oocytes and an ovarian cavity (OC), which exhibits morphological characteristics of the ovary. In *amh* mutants, the genetic male and female gonads showed typically normal ovaries containing many oocytes and an ovarian cavity. Scale bars, 50 μm.

upregulated from 60 dah and showed increasing expression from 65 dah in XX samples (Figure 3D). Localization *amh* mRNAs by *in situ* hybridization showed positive signals in gonads before and during gonadal sex differentiation only in the XY genotype (Figure 3E). At 35 dah, the signals were detected in presumptive Sertoli cells, whereas at 100 dah, when clear characteristics of testis differentiation can be detected, strong signals were observed in presumptive Sertoli cells surrounding cysts of germ cells at the ventral side of the gonad. Although, the expression of *amhx* was detected by qRT-PCR in some samples, the levels were extremely low compared with the expression levels of

TABLE 5 Genotypic and phenotypic sexes in juveniles of *P. olivaceus* and the types and frequencies of *amh* mutations in each individual.

| Genotype | Genotypic sex | Phenotypic sex (gonad) | Number of clones sequenced | Types of mutation | | | Rate of long deletion (%) |
|---|---|---|---|---|---|---|---|
| | | | | Long deletion | Shorts indels | No mutation | |
| Wild-type | XY | ♂ (Testis) | 4 | 0 | 0 | 4 | 0 |
| | XX | ♀ (Ovary) | 4 | 0 | 0 | 4 | 0 |
| *amh* mutant | XY | ♀ (Ovary) | 16 | 13 | 3 | 0 | 81.3 |
| | XY | ♀ (Ovary) | 16 | 12 | 4 | 0 | 75.0 |
| | XY | ♀ (Ovary) | 16 | 12 | 4 | 0 | 75.0 |
| | XX | ♀ (Ovary) | 8 | 6 | 2 | 0 | 75.0 |
| | XX | ♀ (Ovary) | 7 | 5 | 2 | 0 | 71.4 |

*amhy* (Figure 3F), which suggests that the *in situ* hybridization signals detected in XY gonads most probably represent the expression of *amhy*.

Analysis of the phenotypic sex at the end of the rearing experiment (100 dah) by gonadal histology in 18 larvae showed that individuals with differentiating ovaries were all *amhy*-negative XX ($n$ = 8) and those with a differentiating testis were all *amhy*-positive XY ($n$ = 10).

## Generation of *amh*-mutant Japanese flounder using the CRISPR-Cas9 system

To elucidate the roles of a*mh* in Japanese flounder, we produced *amh*-mutant flounder using the CRISPR-Cas9 system. We selected two target sequences to obtain wide-range deletion mutations that deleted the functional regions of *amh* (Sawamura et al., 2017). The crRNA target sequences were designed in exon 7 of the *amh* gene, which encodes a conserved TGF-beta domain at its 3′ end. We expected that by deleting the approximately 680-bp region the encoded Amh protein would lack function (Figure 4A). The two crRNAs were co-injected into fertilized eggs, and hatchlings were reared until 200 dah to investigate the mutation efficiency. The amplicon DNA fragments of the *amh*-mutants were examined by sequence analysis, which showed 682 bp or 684 bp deletions between the two target sequences and some indel mutations in both target sequences (Figure 4A). The wide-range deletion efficiency in genetic male fish was 81.3%, 75.0%, and 75%, and the efficiency in genetic female fish was 75.0% and 71.4% (Table 5).

## Loss of Amh function causes male-to-female sex reversal in *P. olivaceus*

To investigate the phenotypes of the *amh*-mutants in *P. olivaceus*, we performed histological analysis of the gonads of wild-type (control) and *amh*-mutant Japanese flounder at 200 dah. Histological observations of the gonads showed that in the controls, all the genetic male gonads were typical testes containing many spermatogonia (Figure 4B), and all the genetic female gonads were typical ovaries harboring many oocytes and an ovarian cavity. In the *amh*-mutants, all the genetic male and female gonads were typical normal ovaries containing many oocytes and an ovarian cavity. These results indicate that the loss of Amh function caused male-to-female sex reversal in *P. olivaceus*.

# Discussion

The number of genome sequencing projects in teleost fish has increased in recent years, and reports of genetic switches of sex determination have increased for many other fish species (Hattori et al., 2020). In this study, we combined sex-linked microsatellite marker information and genome database analysis, and selected genes related to gonadal differentiation from a list of 181 predicted genes. The detailed comparative sequence analysis of the predicted genes identified a male-specific *amh* gene (named as *amhy*) in the sex-linked locus. Although the coding region of *amhy* was identical to that of *amhx*, except for one synonymous SNP, the presumptive *amhy* promoter differed from that of the female-specific *amhx* due to three deletions within a 1-kb fragment. These deletions are seemingly associated with the early expression of *amhy* in XY, which occurs before the onset of histological sex differentiation. These characteristics place *amhy* as the master sex-determining gene of the Japanese flounder *Paralichthys olivaceus*.

The longest deletion in the presumptive *amhy* promoter corresponds to a region close to the start codon (at position −14 bp) and the other two deletions are located in a further upstream region. Which of these deletions are responsible for the early expression of *amhy* needs further examination by promoter and/or functional analysis. Nevertheless, because the coding regions of *amhy* and *amhx* are almost identical, it seems reasonable to assume that modifications in their regulatory

regions are responsible for the different expression profile of these two genes. Some sex-determining genes reported previously have differences in the coding nucleotide and amino acid sequences, and these differences are associated with genotypic sex determination in *Takifugu rubripes amhrII* (Kamiya et al., 2012) and *Seriola quinqueradiata hsd17b1* (Koyama et al., 2019). The pattern of *amhy* expression in *P. olivaceus* is similar to that of sex-determining genes in the medaka *Oryzias luzonensis sox3*$^{Y}$ (Takehana et al., 2014) that differ by only a 9-bp deletion in the promoter region.

Although many sex-determining genes have been described in teleost fish (Hattori et al., 2020), there seems to be a high likelihood that *amh* will take over the position of sex-determining gene (Bej et al., 2017; Hattori et al., 2019; Pan et al., 2021; Song et al., 2021), and this has been corroborated by this study. Besides the sex linkage, another characteristic shared by *amhy* genes of other species is their high expression in pre-Sertoli cells before the onset of histological sex differentiation, which is seemingly associated with changes in promoter regions. In *P. olivaceus*, we detected the expression of *amhy* from 20 dah with a peak at 25 dah. Although we did not assess the expression before 20 dah, 25 dah can be considered as relatively early (Kitano et al., 1999; Kitano et al., 2000) and much earlier than the upregulation of *cyp19a1a* in XX, which started to increase at 65 dah. The expression of *amhrII*, the receptor that is supposed to bind to *amh*, was also detected at 20 dah (although at low levels) and increased gradually thereafter. Interestingly, the expression of *amhx* was higher in XX than it was in XY genotypes. Although this may be because XX individuals have two copies of *amhx* and XY individuals only one, this expression pattern deserves further investigation considering that *amhx* was upregulated in XX fish during female-to-male sex reversal as was reported previously (Yamaguchi et al., 2010).

The development of ovaries in XY genotypes (i.e., male-to-female sex reversal) by disrupting the region of the *amh* (presumably *amhy*) gene that encodes the TGF-beta domain using CRISPR-Cas9 technique demonstrated that this gene is necessary for testicular formation in *P. olivaceus* genotypic males and indicated that its downstream function may be mediated by the TGF-beta domain. How widespread the *amhy* of *P. olivaceus* is among other closely-related groups needs further investigation, but the *amhy* gene of *P. olivaceus* shares high similarity with the *amhy* genes of other species that have been described so far, which may indicate recent evolution of *amhy* genes. Considering the sex-determining candidates *dmrt1* in tongue sole (Chen et al., 2014) and *sox2* in turbot (Martínez et al., 2021), the discovery of genes other than *amhy* seems to illustrate that genotypic sex determination in flatfishes in order Pleuronectiformes may involve a variety of sex-determining genes.

In conclusion, the present results support the view that *amhy* is the master sex-determining gene of Japanese flounder *Paralichthys olivaceus*. Further promoter analysis studies may help us to understand the importance of this gene and the transcription regulation behind its expression profile. The availability of a reliable marker of genotypic sex will be instrumental for understanding the interactions between genotypic and environmental sex determination from aquaculture and ecological perspectives.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ddbj.nig.ac.jp/, BRVK01000001–BRVK01002790.

## Ethics statement

This study was carried out in accordance with the Guide for the Care and Use of Laboratory Animals from Tokyo University of Marine Science and Technology (TUMSAT). Experiments with fish at TUMSAT do not require any special authorization as long as they adhere to the institutional guidelines, which is the case of this study.

## Author contributions

Conceived and designed the experiments: TS, RSH. Performed the experiments: RSH, KK, YN, TY, and KF. Analyzed the data: RSH, KK, MN, YN, TY, TK, EY, and KF. Wrote the article: RSH, TK, and TS.

## Funding

This work was supported by grants from the Japan Society for the Promotion of Science (grant numbers 25292118 and 16H04970 to TS and L19549 to RSH) and the Japan International Cooperation Agency (JICA) for Science and Technology Research Partnership for Sustainable Development (SATREPS) to TS.

## Acknowledgments

We thank Profs. Keita Suzuki and Reiji Masuda (Maizuru Fisheries Research Station, Kyoto University) for collecting the wild juveniles. We thank Margaret Biswas, PhD, from Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

## Conflict of interest

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

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1007548/full#supplementary-material

## References


Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-2

Baroiller, J. F., and D'Cotta, H. (2016). The reversible sex of gonochoristic fish: Insights and consequences. *Sex. Dev.* 10, 242–266. doi:10.1159/000452362

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., et al. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi:10.1093/nar/gkaa1100

Bej, D. K., Miyoshi, K., Hattori, R. S., Strüssmann, C. A., and Yamamoto, Y. (2017). A duplicated, truncated amh gene is involved in male sex determination in an old world silverside. *G3(Bethesda)* 7, 2489–2495. doi:10.1534/g3.117.042697

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94. doi:10.1006/jmbi.1997.0951

Castañeda Cortés, D. C., Arias Padilla, L. F., Langlois, V. S., Somoza, G. M., and Fernandino, J. I. (2019). The central nervous system acts as a transducer of stress-induced masculinization through corticotropin-releasing hormone B. *Development* 146, dev172866–10. doi:10.1242/dev.172866

Castaño-Sánchez, C., Fuji, K., Ozaki, A., Hasegawa, O., Sakamoto, T., Morishima, K., et al. (2010). A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). *BMC genomics* 11, 554. doi:10.1186/1471-2164-11-554

Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* 46, 253–260. doi:10.1038/ng.2890

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi:10.1093/bioinformatics/btr330

Fernandino, J. I., Hattori, R. S., Moreno Acosta, O. D., Strüssmann, C. A., and Somoza, G. M. (2013). Environmental stress-induced testis differentiation: Androgen as a by-product of cortisol inactivation. *Gen. Comp. Endocrinol.* 192, 36–44. doi:10.1016/j.ygcen.2013.05.024

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

Hattori, R. S., Castañeda-Cortés, D. C., Arias Padilla, L. F., Strobl-Mazzulla, P. H., and Fernandino, J. I. (2020). Activation of stress response axis as a key process in environment-induced sex plasticity in fish. *Cell. Mol. Life Sci.* 77, 4223–4236. doi:10.1007/s00018-020-03532-9

Hattori, R. S., Fernandino, J. I., Kishil, A., Kimura, H., Kinno, T., Oura, M., et al. (2009). Cortisol-induced masculinization: Does thermal stress affect gonadal fate in pejerrey, a teleost fish with temperature-dependent sex determination? *PLoS One* 4, e6548. doi:10.1371/journal.pone.0006548

Hattori, R. S., Murai, Y., Oura, M., Masuda, S., Majhi, S. K., Sakamoto, T., et al. (2012). A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. *Proc. Natl. Acad. Sci. U. S. A.* 109, 2955–2959. doi:10.1073/pnas.1018392109

Hattori, R. S., Somoza, G. M., Fernandino, J. I., Colautti, D. C., Miyoshi, K., Gong, Z., et al. (2019). The duplicated Y-specific amhy gene is conserved and linked to maleness in silversides of the genus Odontesthes. *Genes* 10, 679. doi:10.3390/genes10090679

Hayashi, Y., Kobira, H., Yamaguchi, T., Shiraishi, E., Yazawa, T., Hirai, T., et al. (2010). High temperature causes masculinization of genetically female medaka by elevation of cortisol. *Mol. Reprod. Dev.* 77, 679–686. doi:10.1002/mrd.21203

Honeycutt, J. L., Deck, C. A., Miller, S. C., Severance, M. E., Atkins, E. B., Luckenbach, J. A., et al. (2019). Warmer waters masculinize wild populations of a fish with temperature-dependent sex determination. *Sci. Rep.* 9, 6527–6613. doi:10.1038/s41598-019-42944-x

Ieda, R., Hosoya, S., Tajima, S., Atsumi, K., Kamiya, T., Nozawa, A., et al. (2018). Identification of the sex-determining locus in grass puffer (*Takifugu niphobles*) provides evidence for sex-chromosome turnover in a subset of *Takifugu* species. *PLoS One* 13, e0190635. doi:10.1371/journal.pone.0190635

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., et al. (2012). A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger Pufferfish, *Takifugu rubripes* (Fugu). *PLoS Genet.* 8, e1002798. doi:10.1371/journal.pgen.1002798

Kitano, T., Takamune, K., Kobayashi, T., Nagahama, Y., and Abe, S. I. (1999). Suppression of P450 aromatase gene expression in sex-reversed males produced by rearing genetic female larvae at a high water temperature during a period of sex differentiation in the Japanese flounder (*Paralichthys olivaceus*). *J. Mol. Endocrinol.* 23, 167–176. doi:10.1677/jme.0.0230167

Kitano, T., Takamune, K., Nagahama, Y., and Abe, S. I. (2000). Aromatase inhibitor and 17α-methyltestosterone cause sex-reversal from genetical females to phenotypic males and suppression of P450 aromatase gene expression in Japanese flounder (*Paralichthys olivaceus*). *Mol. Reprod. Dev.* 56, 1–5. doi:10.1002/(SICI)1098-2795(200005)56:1<1::AID-MRD1>3.0.CO;2-3

Koyama, T., Nakamoto, M., Morishima, K., Yamashita, R., Yamashita, T., Sasaki, K., et al. (2019). A SNP in a steroidogenic enzyme is associated with phenotypic sex in Seriola fishes. *Curr. Biol.* 29, 1901–1909.e8. e8. doi:10.1016/j.cub.2019.04.069

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 5, 1754–1760. doi:10.1093/bioinformatics/btp324

Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in nile Tilapia, *Oreochromis niloticus*. *PLoS Genet.* 11, e1005678. doi:10.1371/journal.pgen.1005678

Mankiewicz, J. L., Godwin, J., Holler, B. L., Turner, P. M., Murashige, R., Shamey, R., et al. (2013). Masculinizing effect of background color and cortisol in a flatfish with environmental sex-determination. *Integr. Comp. Biol.* 53, 755–765. doi:10.1093/icb/ict093

Manly, K. F., and Olson, J. M. (1999). Overview of QTL mapping software and introduction to Map Manager QT. *Mamm. Genome* 10, 327–334. doi:10.1007/s003359900997

Martínez, P., Robledo, D., Taboada, X., Blanco, A., Moser, M., Maroso, F., et al. (2021). A genome-wide association study, supported by a new chromosome-level genome assembly, suggests *sox2* as a main driver of the undifferentiatiated ZZ/ZW sex determination of turbot (*Scophthalmus maximus*). *Genomics* 113, 1705–1718. doi:10.1016/j.ygeno.2021.04.007

Miyoshi, K., Hattori, R. S., Strüssmann, C. A., Yokota, M., and Yamamoto, Y. (2020). Phenotypic/genotypic sex mismatches and temperature-dependent sex determination in a wild population of an Old World atherinid, the cobaltcap silverside *Hypoatherina tsurugae*. *Mol. Ecol.* 29, 2349–2358. doi:10.1111/mec.15490

Pan, Q., Feron, R., Jouanno, E., Darras, H., Herpin, A., Koop, B., et al. (2021). The rise and fall of the ancient northern pike master sex determining gene. *Elife* 10, e62858–50. doi:10.7554/eLife.62858

Pan, Q., Feron, R., Yano, A., Guyomard, R., Jouanno, E., Vigouroux, E., et al. (2019). Identification of the master sex determining gene in Northern pike (*Esox lucius*) reveals restricted sex chromosome differentiation. *PLoS Genet.* 15, e1008013. doi:10.1371/journal.pgen.100801310.1101/549527

Peterson, B. C., and David, K. B. (2012). Effects of gender and sex hormones on disease susceptibility of channel catfish to *Edwardsiella ictaluri*. *J. World Aquac. Soc.* 43, 733–738. doi:10.1111/j.1749-7345.2012.00588.x

Sakamoto, T., Danzmann, R. G., Gharbi, K., Howard, P., Ozaki, A., Khoo, S. K., et al. (2000). A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155, 1331–1345. doi:10.1093/genetics/155.3.1331

Sarida, M., Hattori, R. S., Zhang, Y., Yamamoto, Y., and Strüssmann, C. A. (2019). Spatiotemporal correlations between *amh* and *cyp19a1a* transcript expression and apoptosis during gonadal sex differentiation of pejerrey. *Sex. Dev.* 13, 99–108. doi:10.1159/000498997

Sawamura, R., Osafune, N., Murakami, T., Furukawa, F., and Kitano, T. (2017). Generation of biallelic F0 mutants in medaka using the CRISPR/Cas9 system. *Genes Cells.* 22, 756–763. doi:10.1111/gtc.12511

Song, W., Xie, Y., Sun, M., Li, X., Fitzpatrick, C. K., Vaux, F., et al. (2021). A duplicated *amh* is the master sex-determining gene for *Sebastes* rockfish in the Northwest Pacific. *Open Biol.* 11, 11210063210063. doi:10.1098/rsob.210063

Takehana, Y., Matsuda, M., Myosho, T., Suster, M. L., Kawakami, K., Shin-I, T., et al. (2014). Co-option of Sox3 as the male-determining factor on the y chromosome in the fish *Oryzias dancena*. *Nat. Commun.* 5, 4157–4210. doi:10.1038/ncomms5157

Voorrips, R. E. (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi:10.1093/jhered/93.1.77

Yamaguchi, T., Yoshinaga, N., Yazawa, T., Gen, K., and Kitano, T. (2010). Cortisol is involved in temperature-dependent sex determination in the Japanese flounder. *Endocrinology* 151, 3900–3908. doi:10.1210/en.2010-0228

Yamamoto, E. (1999). Studies on sex-manipulation and production of cloned populations in hirame, *Paralichthys olivaceus* (Temminck et Schlegel). *Aquaculture* 173, 235–246. doi:10.1016/s0044-8486(98)00448-7

Yamamoto, E. (1995). Studies on sex-manipulation and production of cloned populations in hirame flounder, *Paralichthys olivaceus* (Temminek et Schlegel). *Bull. Tottori Pref. Fish. Exp. Stn.* 34, 1–145.

Yamamoto, Y., Zhang, Y., Sarida, M., Hattori, R. S., and Strüssmann, C. A. (2014). Coexistence of genotypic and temperature-dependent sex determination in pejerrey Odontesthes bonariensis. *PLoS One* 9, e102574. doi:10.1371/journal.pone.0102574

Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., et al. (2013). The sexually dimorphic on the Y-chromosome gene (*sdY*) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evol. Appl.* 6, 486–496. doi:10.1111/eva.12032

Yoshinaga, N., Shiraishi, E., Yamamoto, T., Iguchi, T., Abe, S. I., and Kitano, T. (2004). Sexually dimorphic expression of a teleost homologue of Müllerian inhibiting substance during gonadal sex differentiation in Japanese flounder, *Paralichthys olivaceus*. *Biochem. Biophys. Res. Commun.* 322, 508–513. doi:10.1016/j.bbrc.2004.07.162

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 03 October 2022
DOI 10.3389/fgene.2022.1002346



OPEN ACCESS

EDITED BY
Nguyen Hong Nguyen,
University of the Sunshine Coast, Australia

REVIEWED BY
Gabriel Campos Montes,
Autonomous Metropolitan University Xochimilco Campus, Mexico
Tengfei Zhu,
Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences (CAFS), China

*CORRESPONDENCE
Md. Mehedi Hasan,
md.m.hasan@sydney.edu.au

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

RECEIVED 25 July 2022
ACCEPTED 15 September 2022
PUBLISHED 03 October 2022

CITATION
Hasan MM, Raadsma HW, Thomson PC, Wade NM, Jerry DR and Khatkar MS (2022), Genetic parameters of color phenotypes of black tiger shrimp (Penaeus monodon). *Front. Genet.* 13:1002346. doi: 10.3389/fgene.2022.1002346



# Genetic parameters of color phenotypes of black tiger shrimp (*Penaeus monodon*)

Md. Mehedi Hasan[1,2]*, Herman W. Raadsma[1,2], Peter C. Thomson[1,2], Nicholas M. Wade[3], Dean R Jerry[2,4] and Mehar S. Khatkar[1,2]

[1]The University of Sydney, Faculty of Science, Sydney School of Veterinary Science, Camden, NSW, Australia, [2]ARC Research Hub for Advanced Prawn Breeding, Townsville, QLD, Australia, [3]CSIRO Agriculture and Food, Queensland Biosciences Precinct, St Lucia, QLD, Australia, [4]Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, Townsville, QLD, Australia

Black tiger shrimp (*Penaeus monodon*) is the second most important aquaculture species of shrimp in the world. In addition to growth traits, uncooked and cooked body color of shrimp are traits of significance for profitability and consumer acceptance. This study investigated for the first time, the phenotypic and genetic variances and relationships for body weight and body color traits, obtained from image analyses of 838 shrimp, representing the progeny from 55 sires and 52 dams. The color of uncooked shrimp was subjectively scored on a scale from 1 to 4, with "1" being the lightest/pale color and "4" being the darkest color. For cooked shrimp color, shrimp were graded firstly by subjective scoring using a commercial grading score card, where the score ranged from 1 to 12 representing light to deep coloration which was subsequently found to not be sufficiently reliable with poor repeatability of measurement ($r$ = 0.68–0.78) Therefore, all images of cooked color were regraded on a three-point scale from brightest and lightest colored cooked shrimp, to darkest and most color-intense, with a high repeatability ($r$ = 0.80–0.92). Objective color of both cooked and uncooked color was obtained by measurement of RGB intensities (values range from 0 to 255) for each pixel from each shrimp. Using the "convertColor" function in "R", the RGB values were converted to *L*a*b** (CIE Lab) systems of color properties. This system of color space was established in 1976, by the International Commission of Illumination (CIE) where "*L**" represents the measure of degree of lightness, values range from 0 to 100, where 0 = pure black and 100 = pure white. The value "*a**" represents red to green coloration, where a positive value represents the color progression towards red and a negative value towards green. The value "*b**" represents blue to yellow coloration, where a positive value refers to more yellowish and negative towards the blue coloration. In total, eight color-related traits were investigated. An ordinal mixed (threshold) model was adopted for manually (subjectively) scored color phenotypes, whereas all other traits were analyzed by linear mixed models using ASReml software to derive variance components and estimated breeding values (EBVs). Moderate to low heritability estimates (0.05–0.35) were obtained for body color traits. For subjectively scored cooked and uncooked color, EBV-based selection would result in substantial genetic improvement in these traits. The genetic

correlations among cooked, uncooked and body weight traits were high and ranged from −0.88 to 0.81. These suggest for the first time that 1) cooked color can be improved indirectly by genetic selection based on color of uncooked/live shrimp, and 2) intensity of coloration is positively correlated with body weight traits and hence selection for body weight will also improve color traits in this population.

KEYWORDS

*Penaeus monodon*, cooked, uncooked, color, genetic parameters

## Introduction

The black tiger shrimp (*Penaeus monodon*) is regarded as a luxury food commodity because of its elegant sensory properties and high nutritional value. For black tiger shrimp, henceforth "shrimp", visual sensory properties of bright red color after cooking contribute to higher prices and greater consumer acceptance (Tume et al., 2009). These color properties are derived from carotenoids (e.g., astaxanthins). Carotenoids are primarily produced by the primary producers (e.g., photosynthetic plants and algal species) of the ecosystems, and like most other secondary consumers, shrimp obtain their carotenoid elements through their diet. The body color of shrimp is mainly regulated by astaxanthin, which mainly presents in the exoskeleton and in the surface of abdominal muscle beneath the exoskeleton (Menasveta et al., 1993; Boonyaratpalin et al., 2001; Tume et al., 2009). Astaxanthin remains in both free and esterified forms with fatty acids. Moreover, when carotenoid astaxanthin binds with protein, it can also be found as carotenoproteins. In marine invertebrates such proteins cause a big change in the carotenoid light absorption spectrum to produce a range of bright coloration, e.g., purple, blue, red or green (Britton et al., 1982). This bright coloration becomes more apparent when the shrimp are cooked, since cooking causes the carotenoprotein complexes to dissociate, resulting in increased color brightness, e.g., the typical bright red coloration of cooked shrimp (Britton et al., 1982; Tume et al., 2009; Wade and Glencross, 2011).

Besides coloration, carotenoid astaxanthin has a role in various other important physiological functions in shrimp, including growth, reproductive competency, survival, disease, and stress resistance (Supamattaya et al., 2005; Paibulkichakul et al., 2008; Niu et al., 2014; Wade et al., 2017). A study by Niu et al. (2014) has shown that diet supplementation of astaxanthin significantly enhanced the growth and immunological competency in *P. monodon*. Similarly, shrimp fed with blue green algae (*Dunaliella* sp.) containing astaxanthin, resulted in higher weight gain, survival, resistance to white spot syndrome virus (WSSV) infection and greater tolerance to stress conditions (e.g., low dissolved oxygen) (Supamattaya et al., 2005). An increase in egg numbers and spermatozoa was associated with elevated levels of astaxanthin in *P. monodon* (Paibulkichakul et al., 2008). An extensive review of the role and function of carotenoids in crustacean aquaculture revealed that carotenoids are essential for overall growth, performance, and coloration in shrimp (Wade et al., 2017). Since the source of carotenoid astaxanthin for shrimp comes from dietary sources, e.g., natural algae or added carotenoid astaxanthin in the dietary pellets, the addition of pigmentation additives causes a substantial increase in production costs. Numerous studies have shown that there is both an environmental and a genetic component of pigmentation for shrimp and aquaculture species. Background color of the rearing tank/environment tends to cause significant variation in animal pigmentation, e.g., *P. monodon* reared in a black tank display darker coloration (Wade et al., 2012). The genetic basis for pigmentation has been confirmed for several aquaculture species (e.g., salmon, rainbow trout), where some individuals are genetically superior in absorbing, transporting and depositing carotenoid astaxanthin from feed. However, no such information is available for *P. monodon*.

*P. monodon* is currently the major aquaculture crustacean species farmed in Australia and economically, is the second most important species in the world with an economic value of USD 5.7 billion and a production base of 750,600 tons in 2018 (FAO, 2020). Studies have shown that its current economic value can be increased significantly by improving color phenotypes. For example, AU$2–4/kg can be added for dark red colored shrimp over pale colored ones (Tume et al., 2009). To study and implement body color traits in selective breeding programs, the phenotyping of the traits under selection must be straightforward, standardised, relatively low cost, and accurate. To date there is no established protocol for measuring body color phenotypes for *P. monodon*. Measuring body color phenotype is complex, with patterned banding and uneven distribution of pigmentation, when compared to other traits (e.g., body weight). Its measurement can be influenced by various external factors, e.g., frequent changing of lighting conditions can affect the visual assessment of the body color of the shrimp. Broadly, there are two means for color phenotyping, namely 1) chemical, and 2) physical measurement approaches. In the chemical measurement approach, the phenotyping is done indirectly by NIR (near-infrared reflectance) or directly through HPLC analysis by quantifying color-producing chemical components in the shrimp. However, chemical analysis is expensive and time

consuming and requires destructive sampling of animals, although it provides an accurate measurement of color phenotypes. On the other hand, the physical approach involves either visual determination of color intensity using a standardized color card, or by using an instrumental measurement where a tristimulus colorimeter (chromometer) measures the reflectance of light from the subject (e.g., shrimp) compared to a background calibration plate. Color may then be expressed in the $L^*a^*b^*$ system, with $L^*$ measuring lightness, $a^*$ redness/greenness and $b^*$ yellowness/blueness (Robertson, 1977). Studies have shown that this instrumental measurement approach is potentially very useful in color phenotype studies in aquaculture species, as the value $a^*$ (intensity of redness) is linearly correlated with the carotenoid pigment content in salmon fish flesh (Christiansen et al., 1995).

Although the body color phenotype has direct economic benefit to the shrimp aquaculture industry, no study has been conducted so far for genetic improvement of this trait in this species. Establishing the genetic basis for variation in pigmentation in *P. monodon*, will help to identify brood stock with superior color phenotypes. This will reduce the need for adding dietary astaxanthin level and will ultimately increase the overall profitability in *P. monodon* aquaculture (De Carvalho and Caramujo, 2017). Moreover, it is unknown whether selection on body color traits would have any detrimental effect on other commercially important traits (e.g., body weight, body length). Furthermore, it is unknown whether uncooked shrimp color is genetically correlated to cooked color. This information is particularly important, i.e., if there is any positive association between these two traits, then there will be opportunity to select live/uncooked body color of shrimp to improve the cooked color phenotype. The aims of the study were therefore: 1) To evaluate the efficiency of subjective measurement of color phenotypes based on manual scoring and compare with instrumental color analysis in *P. monodon*, 2) to estimate the genetic basis of the pigmentation phenotypes, 3) assess genetic correlation among cooked, uncooked color and other economically-important morphological traits and finally, 4) compare genetic gains by direct and indirect selection in color phenotypes.

# Material and methods

## Origin of the study population

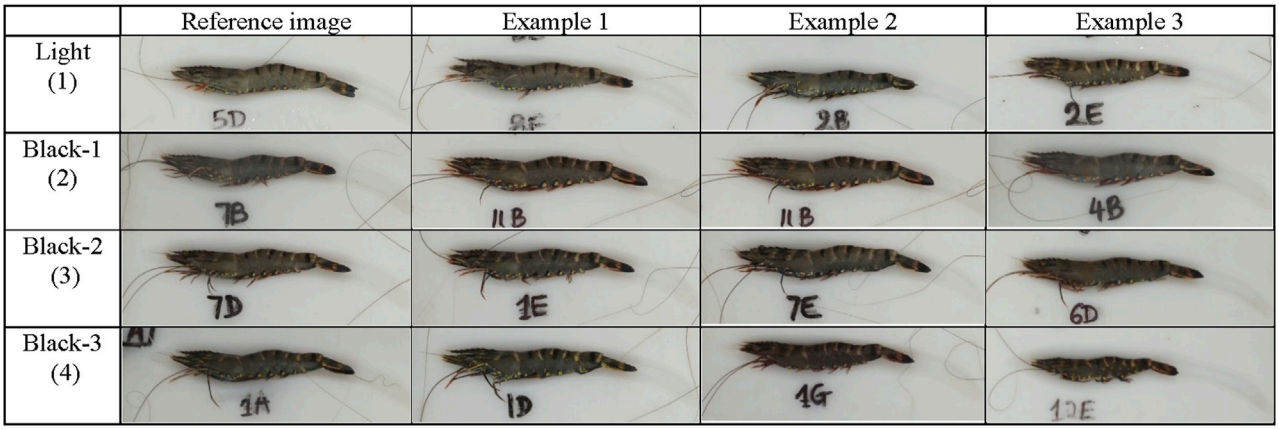
All progeny in this experiment were sampled from commercial cohorts of *P. monodon* raised by Seafarms Group Ltd., as described in Foote et al. (2019). Briefly, wild broodstock were sourced from the Northern Territory, Australia and transferred to a commercial hatchery at Flying Fish Point, Queensland, Australia. Broodstock maturation was conducted within indoor flow-through tank systems (density of 3 $m^{-2}$ at 28°C ± 0.5°C) and broodstock were fed a commercial maturation diet. For each cohort, broodstock were allowed to mate naturally within the tanks, with any unmated females then artificially inseminated following industry practices. Since tracing of broodstock contribution could not be done on farm, all potential broodstock were genotyped and parentage analysis was utilised to determine the contributing parents retrospectively as detailed by Guppy et al. (2020). Females were spawned in communal spawning tanks and spawned eggs were transferred hatching tanks, and hatched nauplii were then transferred into 20,000 L larval rearing tanks (LRTs) at a density from 100 to 125 nauplii/liter, and reared on a commercial diet until 30 DOC. LRTs were then pooled and stocked into 4,000 $m^2$ grow-out ponds and reared under commercial conditions at a density of 45 $m^{-1}$ until harvest. Immediately pre-harvest ponds were sampled by random castnet. The current study population comprised of 55 sires and 52 dams. In total 67 full sib and half-sib families were produced across 838 progeny and stocked across seven ponds as shown in Supplementary Table S1. From post-larval stage to harvest, the growth periods ranged from 124 to 143 days across ponds. Throughout the grow-out period the key water quality parameters were recorded, including dissolved oxygen, temperature, pH and salinity (pond water quality parameter has been provided in Hasan (2022).

The genotyping method described by Noble et al. (2020) was employed to determine pedigree structure. A genotype-by-sequence (GBS) based single nucleotide polymorphism (SNP) genotyping method was used for the brood stocks (DArTSeq) (Sansaloni et al., 2011). This DArTSeq data set was used to derive a targeted 4 K DArTcap custom SNP panel (4,194 SNPs) for genotyping of the offspring (Guppy et al., 2020). CERVUS version 3.0.7 (Kalinowski et al., 2007) was used to perform family assignment, and Colony V2.0.6.4 (Jones and Wang, 2010) was employed to allocate offspring to the appropriate genetic group. When the parental information was missing, an arbitrary parental ID was given to each group.

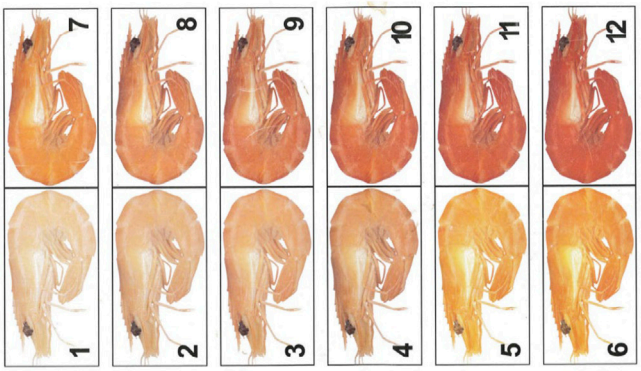
## Phenotypic recording and characterization for genetic analysis

The color of uncooked shrimp was subjectively scored by a single individual on a scale from 1 to 4, with "1" being the lightest/pale color and "4" being the darkest color (Figure 1A). For cooked color of the same shrimp, the shrimp were graded firstly by subjective scoring using a commercial grading score card, where the score ranged from 1 to 12 representing light to deep coloration (Aqua-marine Marketing Pty. Ltd., Kippa-Ring, Queensland, Australia) (Figure 1B). However, the repeatability (Pearson correlation coefficient) of color scoring was not considered sufficiently reliable using the commercial grading score card ($r$ = 0.72) (Table 3). Therefore, by manual

A




B




C

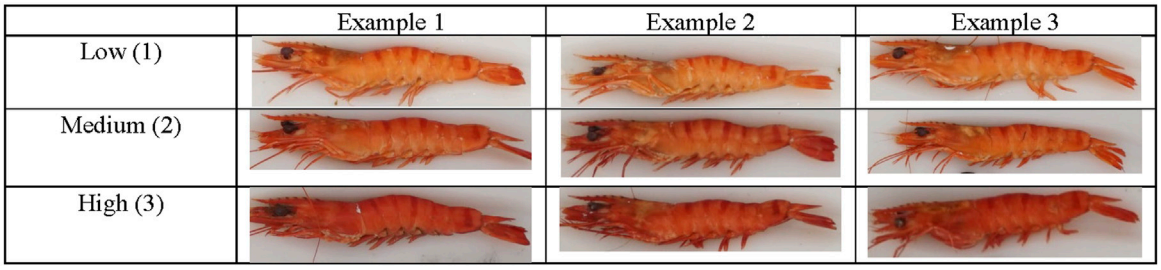



FIGURE 1
The reference image used for the scoring of uncooked and cooked shrimp. **(A)** Uncooked shrimp color chart, **(B)** commercial grade shrimp color score card (source: Aqua-Marine Marketing, Newport, QLD) and **(C)** cooked shrimp color chart.

inspection of the shrimp images from brightest and lightest colored to the darkest and most color-intense cooked shrimp, three grades of colors (a scale of 1–3) were identified and selected as reference images for color grading of cooked shrimp for scoring (Figure 1C). The reliability of this new cooked color shrimp scoring was evaluated by examining confusion matrix tables and estimating repeatability of the manual scoring on a sample of 288 images (Table 3).

Objective color was measured by a standard digital photographs. The instrument returned RGB intensities (values range from 0 to 255) for each pixel from each shrimp sampled. Using the "convertColor" function in "R", the RGB values were converted to $L^*a^*b^*$ (CIE Lab) systems of color properties. The CIE Lab color assessment system aligns more closely with human perception of color (Gómez-Polo et al., 2016; Ly et al., 2020). This system of color space was established in 1976, by the

International Commission of Illumination (CIE). Here the "$L^*$" represents the measure of degree of lightness, values range from 0 to 100, where 0 = pure black and 100 = pure white. The value "$a^*$" represents red to green coloration, where a positive value represents the color progression towards red and a negative value towards green. The value "$b^*$" represents blue to yellow coloration, where a positive value refers to more yellowish and negative towards the blue coloration.

## Statistical analysis

The manually scored color traits were considered as ordinal categorical variables and an ordinal logistic mixed model was employed to estimate variance components. This model considers three and four scores from each of cooked and uncooked shrimp, respectively. For each single observation in the data set, the model has the following form:Where $Y_{ij}$ is the color score of the $i$th shrimp in the $j$th pond, and $k$ is the (ordinal) score threshold, with $c$ being the number of score points ($c$ = 3 points for cooked color and $c = 4$ points for uncooked colors. $\theta_k$ is the intercept for each of the score points. $\text{Pond}_j$ is a fixed effect and $a_i$ is the polygenic random effect of the individual shrimp, linked to the pedigree, with $\mathbf{a} = \{a_i\}$ and where $\mathbf{a} \sim N(\mathbf{0}, \sigma_A^2\mathbf{A})$ where $\mathbf{A}$ is the numerator relationship matrix. This type of ordinal logistic regression is also known as the proportional odds model (Agresti, 2003). Heritability was calculated on a liability scale as follows:

$$h^2 = \sigma_A^2 / (\sigma_A^2 + \pi^2/3)$$

where $\pi^2/3$ is the liability residual variance and $\sigma_A^2$ is the variance estimate attributed to the additive genetic effects.

$$\log_e[\mathrm{P}(Y_{ij} \le k)/\mathrm{P}(Y_{ij} > k)] = \theta_k + \text{Pond}_j + a_i\,,\ k = 1, \ldots, c{-}1$$

The following linear mixed model was employed for estimating variance components for each of the growth traits and body color traits (i.e., $L^*_{\text{uncooked}}$, $a^*_{\text{uncooked}}$, $b^*_{\text{uncooked}}$, $L^*_{\text{cooked}}$, $a^*_{\text{cooked}}$ and $b^*_{\text{cooked}}$):

$$y_{ij} = \mu + \text{Pond}_j + a_i + \varepsilon_{ij}$$

where $y_{ij}$ is the observation of individual $i$ in pond $j$, μ is the mean and $\text{Pond}_j$ is the fixed effect of the $j$th pond, $a_i$ is the additive genetic effect, both terms as defined in the ordinal model, and $\varepsilon_{ij}$ is the random error, assumed $N(0, \sigma_e^2)$. Heritability was estimated as

$$h^2 = \sigma_A^2 / (\sigma_A^2 + \sigma_e^2)$$

Genetic and phenotypic correlations among the traits studied were estimated using bivariate mixed models, of the form

$$\begin{pmatrix} y_{ij1} \\ y_{ij2} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} \text{Pond}_{j1} \\ \text{Pond}_{j2} \end{pmatrix} + \begin{pmatrix} a_{i1} \\ a_{i2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{ij1} \\ \varepsilon_{ij2} \end{pmatrix}$$

TABLE 1 Phenotypic means, standard deviations and co-efficient of variation of body weight, body length and objectively measured color related traits of shrimp before and after cooking.

| Trait | *n* | Mean | SD | CV% | Min | Max |
|---|---|---|---|---|---|---|
| Body weight (g) | 838 | 13.64 | 3.50 | 25.7 | 1.04 | 26.21 |
| Body length (cm) | 838 | 10.56 | 0.95 | 9.0 | 4.78 | 12.85 |
| $L^*_{\text{uncooked}}$ | 838 | 23.84 | 3.82 | 16.0 | 14.96 | 37.71 |
| $a^*_{\text{uncooked}}$ | 838 | 0.80 | 1.72 | 209 | −3.95 | 6.64 |
| $b^*_{\text{uncooked}}$ | 838 | 9.80 | 2.04 | 20.8 | 3.73 | 16.01 |
| $L^*_{\text{cooked}}$ | 838 | 44.06 | 4.12 | 9.4 | 30.72 | 58.13 |
| $a^*_{\text{cooked}}$ | 838 | 56.85 | 6.00 | 10.6 | 17.01 | 70.72 |
| $b^*_{\text{cooked}}$ | 838 | 55.84 | 6.46 | 11.6 | 15.45 | 70.67 |

with terms defined as in the above univariate model, and subscripts 1 and 2 indicating the pair of traits 1 and 2. As well as variance component estimates as outlined for the univariate models, covariances between trait pairs for additive genetic and residual effects were estimated. Phenotypic and genetic correlation estimates were obtained using these variance and covariance component estimates. As software is not available for bivariate ordinal-ordinal and linear-ordinal models, to estimate the genetic correlation among ordinal and numerical traits, Pearson correlations between the estimated breeding values (EBVs) were calculated (Calo et al., 1973).

Indirect genetic selection, i.e. correlated response in trait $y$ with 1 standard deviation (SD) selection differential in trait $x$, was calculated from the following equation (Falconer and Mackay, 1996):

$$CR_y = r_g \times h_x \times h_y \times SD_y$$

where $SD_y$ is the SD of trait "$y$". The correlated response in trait "$y$" as a percentage of gain possible from direct selection for trait "$x$" is calculated as % of indirect selection (IS), the relative efficiency of correlated response (CR) in trait $y$ when selection is applied on trait $x$ as a percentage of gain possible from direct selection for trait $y$, i.e.,

$$\%IS = \frac{CR_y}{SD_y} \times 100 = r_g \times h_x \Big/ h_y \times 100$$

Data analysis was performed in R v 4.1.0 (R Core Team, 2021), and the genetic analyses using the ordinal and linear mixed models (variance/covariance estimation, EBV calculation) were performed using ASReml-R 4.0 (VSNi) (Butler et al., 2017). Note that, the estimated breeding values (EBVs) are taken as the best linear unbiased predictions (BLUPs) of the $a_i$ in both ordinal and linear models.

TABLE 2 Distribution of manual scores for cooked and uncooked shrimp from 838 animals.

| Trait | Appearance scores | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Uncooked | Light $n$ = 29 (3.5%) | Black-1 $n$ = 272 (32.5%) | Black-2 $n$ = 424 (50.6%) | Black-3 $n$ = 113 (13.5%) |
| Cooked | Light orange $n$ = 92 (11.0%) | Medium orange $n$ = 472 (56.3%) | Bright orange $n$ = 274 (32.7%) | — |

TABLE 3 Confusion matrix and repeatability of repeated measurement of cooked shrimp based on commercial color chart and the one derived for this study. The repeatability is indicated by the correlation ($r$) between the replicates (1 vs. 2, 1 vs. 3 and 2 vs. 3) for both color-scoring systems.

| Commercial system scores | | | | | |
|---|---|---|---|---|---|
| **Replicate 2** | | | | | |
| | | **Score = 9** | **Score = 10** | **Score = 11** | |
| Replicate 1 | Score = 9 | 60 | 10 | 0 | $r$ = 0.78 |
| | Score = 10 | 5 | 12 | 1 | |
| | Score = 11 | 0 | 2 | 4 | |
| Replicate 3 | | | | | |
| | | Score = 9 | Score = 10 | Score = 11 | |
| Replicate 1 | Score = 9 | 53 | 12 | 0 | $r$ = 0.71 |
| | Score = 10 | 6 | 16 | 2 | |
| | Score = 11 | 0 | 3 | 2 | |
| Replicate 3 | | | | | |
| | | Score = 9 | Score = 10 | Score = 11 | |
| Replicate 2 | Score = 9 | 54 | 16 | 0 | $r$ = 0.68 |
| | Score = 10 | 5 | 12 | 1 | |
| | Score = 11 | 0 | 3 | 3 | |
| cores derived in this study | | | | | |
| Replicate 2 | | | | | |
| | | Score = 1 | Score = 2 | Score = 3 | |
| Replicate 1 | Score = 1 | 18 | 6 | 0 | $r$ = 0.81 |
| | Score = 2 | 7 | 82 | 1 | |
| | Score = 3 | 0 | 18 | 56 | |
| Replicate 3 | | | | | |
| | | Score = 1 | Score = 2 | Score = 3 | |
| Replicate 1 | Score = 1 | 20 | 5 | 0 | $r$ = 0.92 |
| | Score = 2 | 3 | 101 | 2 | |
| | Score = 3 | 0 | 2 | 55 | |
| Replicate 3 | | | | | |
| | | Score = 1 | Score = 2 | Score = 3 | |
| Replicate 2 | Score = 1 | 18 | 6 | 0 | $r$ = 0.80 |
| | Score = 2 | 5 | 83 | 2 | |
| | Score = 3 | 0 | 19 | 55 | |

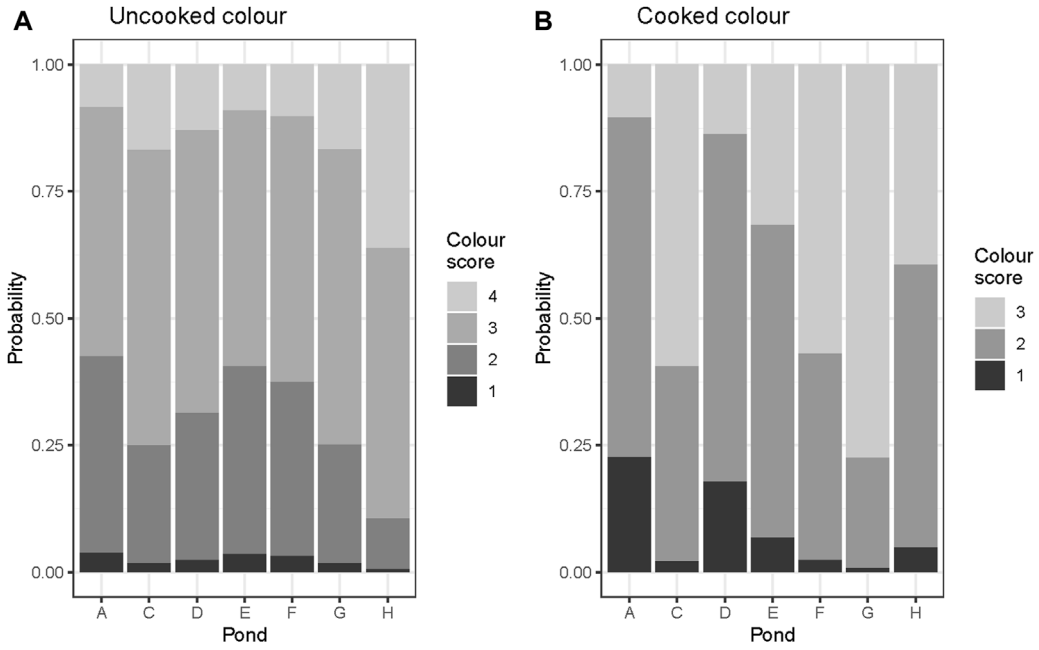



FIGURE 2
Probablility distribution of manual color scores by ponds **(A–G)**, **(A)** uncooked, and **(B)** cooked shrimp.

## Results

The mean, standard deviation, and coefficient of variation for body color, body weight and body length traits of shrimp at harvest are provided in Table 1. The average estimates of the traits studied are provided in Table 1. The description and distribution of the manually-scored color traits are provided in Table 2.

As shown in Table 3, the confusion matrix table revealed that the commercial image scale for grading shrimp body color is not suitable, as suggested by low repeatability ($r$ = 0.72) of repeated scoring on a sample of 288 images. Instead, using images from current population-specific samples, provided more accurate estimates. This was supported by observing the repeatability analysis of color scoring. The repeatability analysis of scoring using the commercial shrimp cooked color chart and using the reference image developed in this study revealed that the scoring system developed in this study is more reliable (repeatability, $r$ = 0.84) than the commercial image scale (repeatability, $r$ = 0.72).

The observation of the probability distribution of color scores across ponds revealed that the bright cooked color shrimp were more common in ponds 'C', 'F' and 'G'. In contrast, medium orange colored shrimp were abundant in ponds 'A', 'D' and 'E' (Figure 2B).

Heritability estimates for manually-scored uncooked and cooked body color traits were both 0.12 ± 0.04. Using instrumental measurement of body color phenotypes, heritability of the $b^*$ trait for uncooked color was highest at 0.35 ± 0.08, followed by $a^*$ trait for uncooked color at 0.29 ± 0.08, whereas they were generally low for cooked color indices. Details of these estimates are presented in Table 4.

TABLE 4 Genetic parameter estimates (±s.e.) for body color and body size traits.

| Traits | $h^2$ | $\sigma_A^2$ | $\sigma_e^2$ |
|---|---|---|---|
| $\text{Manual}_{\text{uncooked}}$ | 0.12 ± 0.04 | 0.48 ± 0.19 | $\pi^2/3$ |
| $\text{Manual}_{\text{cooked}}$ | 0.12 ± 0.04 | 0.46 ± 0.19 | $\pi^2/3$ |
| $L^*_{\text{uncooked}}$ | 0.14 ± 0.05 | 1.94 ± 0.83 | 11.44 ± 0.79 |
| $a^*_{\text{uncooked}}$ | 0.29 ± 0.08 | 0.72 ± 0.22 | 1.71 ± 0.16 |
| $b^*_{\text{uncooked}}$ | 0.35 ± 0.08 | 1.51 ± 0.44 | 2.75 ± 0.29 |
| $L^*_{\text{cooked}}$ | 0.08 ± 0.04 | 1.06 ± 0.53 | 10.94 ± 0.66 |
| $a^*_{\text{cooked}}$ | 0.05 ± 0.03 | 1.78 ± 1.17 | 30.92 ± 1.76 |
| $b^*_{\text{cooked}}$ | 0.06 ± 0.03 | 2.03 ± 1.26 | 29.19 ± 1.72 |
| Body weight | 0.27 ± 0.07 | 3.95 ± 1.25 | 10.20 ± 0.91 |
| Body length | 0.32 ± 0.08 | 0.34 ± 0.10 | 0.69 ± 0.07 |

$h^2$ = heritability, $\sigma_A^2$ = additive genetic variance, $\sigma^2_e$ = environmental variance.

Strong genetic correlations were observed among several cooked and uncooked color traits (ranged from 0.82 to −0.88). Of these, strong positive genetic correlations were seen among $L^*_{\text{uncooked}}$ : $L^*_{\text{cooked}}$ ($r$ = 0.82), similarly uncooked color score had a strong genetic correlation with cooked color score ($r$ = 0.77). Of note is that

TABLE 5 The genetic (lower diagonal) and phenotypic correlations (upper diagonal) among the studied traits.

| | Manual (uncooked) | Manual (cooked) | $L^*_{uncooked}$ | $a^*_{uncooked}$ | $b^*_{uncooked}$ | $L^*_{cooked}$ | $a^*_{cooked}$ | $b^*_{cooked}$ | Body weight | Body length |
|---|---|---|---|---|---|---|---|---|---|---|
| Manual (uncooked) | | 0.44 | −0.74 | 0.32 | −0.28 | −0.47 | 0.36 | −0.04 | −0.12 | −0.12 |
| Manual (cooked) | 0.77 | | −0.57 | 0.28 | −0.30 | −0.62 | 0.59 | −0.18 | −0.03 | −0.03 |
| $L^*_{uncooked}$ | −0.88 | −0.81 | | −0.42 | 0.38 | 0.55 | −0.53 | 0.00 | 0.05 | 0.06 |
| $a^*_{uncooked}$ | 0.33 | 0.36 | −0.25 | | 0.09 | −0.28 | 0.31 | −0.04 | −0.02 | −0.04 |
| $b^*_{uncooked}$ | −0.41 | −0.54 | 0.6 | 0.07 | | 0.24 | −0.20 | 0.25 | 0.11 | 0.11 |
| $L^*_{cooked}$ | −0.82 | −0.84 | 0.82 | −0.37 | 0.43 | | −0.51 | 0.33 | 0.15 | 0.16 |
| $a^*_{cooked}$ | 0.74 | 0.81 | −0.81 | 0.1 | −0.53 | −0.81 | | 0.44 | −0.04 | −0.01 |
| $b^*_{cooked}$ | −0.35 | −0.50 | 0.28 | −0.63 | 0.30 | 0.48 | −0.14 | | 0.15 | 0.20 |
| Body Weight | −0.28 | −0.30 | 0.31 | −0.21 | 0.15 | 0.43 | −0.39 | 0.33 | | 0.93 |
| Body Length | −0.21 | −0.2 | 0.24 | −0.22 | 0.09 | 0.34 | −0.27 | 0.32 | 0.96 | |

$L^*_{uncooked}$ was also strongly correlated with both uncooked and cooked color ($r$ = -0.88 and -0.81, respectively) suggesting this $L^*_{uncooked}$ could be a potential indirect selection marker for both cooked and uncooked color. Similarly, both $L^*_{cooked}$ and $a^*_{cooked}$ had strong genetic correlations with both cooked and uncooked color scores (Table 5), A medium strength genetic correlation was observed between body color and body weight traits (ranged, $r$ = −0.39–0.43) as shown in Table 5. As expected, a correlation of near unity was observed for body weight and body length for both phenotypic ($r$ = 0.93) and genetic ($r$ = 0.96) correlations (Table 5).

The computation of correlated response among the studied traits revealed that the selection on uncooked color with 1 SD selection intensity will lead to a correlated response of 77% on cooked color trait (Table 6). Similarly, selection in the uncooked color score will improve $L^*_{cooked}$ and $a^*_{cooked}$ values with an efficiency of 100% and 114% respectively. Likewise, selection on the $L^*_{uncooked}$ trait will lead to a response of 108.5% on the $L^*_{cooked}$ color trait and a 135% in the $a^*_{cooked}$ values (Table 6). Moderate indirect selection responses (range 13%–90%) were seen in all color traits, both uncooked and cooked, when selection for increased body weight is applied.

Figure 2 shows the distribution of the color scores obtained after fitting all other effects in the model. Distribution of uncooked color (Figure 2A) was more evenly distributed across ponds than cooked color (Figure 2B) suggesting a potential strong local environmental effect on cooked color.

## Discussion

Shrimp color is an economically important trait, associated with consumer acceptability, and must be accurately and inexpensively phenotyped for genetic evaluation (Blay et al., 2021). In most market scenarios, the price of shrimp is based on the color, both for cooked and uncooked states (Peshanoff and Jaensch, 2009; Wade et al., 2017). For cooked shrimp, bright colored individuals have the greatest market demand (Peshanoff and Jaensch, 2009). On the other hand, for uncooked shrimp, either pale or darker animals are preferred, depending on market demand based on socio-economic and cultural demographics (Mehar et al., 2020).

In our study we evaluated the cooked color phenotype of shrimp using both a commercially-used color chart and our own population-specific reference color chart. We found that the repeatability of color scoring is lower when using the commercial shrimp color chart. This may be due to the fact that the commonly available color scale may not match with our studied shrimp population. This suggests that the commercial shrimp color scoring chart may not be extensively applicable for color phenotyping of all the shrimp populations. For raw/uncooked shrimp color we found no standard reference color score chart for this species and like cooked color, we developed our own population-specific color chart for phenotyping. Overall, our findings suggest that, for subjective scoring of shrimp color phenotype, a population-specific reference color chart should be used for more accurate color scoring.

The genetic analysis of the body color traits revealed that there is sufficient genetic variation in these traits to include in a selective breeding program to improve these traits. There are several previous studies that have been conducted for determining heritability estimates of body color in shrimp species (Nguyen et al., 2014; Giang et al., 2019), however to the best of our knowledge, the present study is the first approach to estimate heritability of body color in *P. monodon*.

Heritability for manually scored body color traits were low (0.12 ± 0.04) for both uncooked and cooked color. These findings agreed well with studies with other aquaculture species for color traits (Rye and Gjerde, 1996; Kause et al., 2002). The low heritability estimates in our study for manually-scored color traits, may in part be attributed to the method of scoring. Subjective phenotyping scoring of color traits are subject to be influenced by the person who records it and measuring conditions (e.g., device used), and this can reduce the

TABLE 6 Expected correlated response among color scores and color traits and body weight in both uncooked and cooked responses.

| Trait 1 | Trait 2 | $h^2$ of trait 1 | $h^2$ of trait 2 | $r_g$ | Correlated response | IS % |
|---|---|---|---|---|---|---|
| $Manual_{uncooked}$ | $Manual_{cooked}$ | 0.12 | 0.12 | 0.77 | 0.09 | 77.00 |
| $Manual_{uncooked}$ | $a_u{}^*{}_{ncooked}$ | 0.12 | 0.29 | 0.33 | 0.06 | 21.23 |
| $Manual_{uncooked}$ | $b^*_{uncooked}$ | 0.12 | 0.35 | −0.41 | −0.08 | −24.01 |
| $Manual_{uncooked}$ | $L^*_{cooked}$ | 0.12 | 0.08 | −0.82 | −0.08 | −100.43 |
| $Manual_{uncooked}$ | $a^*_{cooked}$ | 0.12 | 0.05 | 0.74 | 0.06 | 114.64 |
| $Manual_{uncooked}$ | $b^*_{cooked}$ | 0.12 | 0.06 | −0.35 | −0.03 | −49.50 |
| $Manual_{cooked}$ | $L^*_{uncooked}$ | 0.12 | 0.14 | −0.81 | −0.10 | −74.99 |
| $Manual_{cooked}$ | $a^*_{uncooked}$ | 0.12 | 0.29 | 0.36 | 0.07 | 23.16 |
| $Manual_{cooked}$ | $b^*_{uncooked}$ | 0.12 | 0.35 | −0.54 | −0.11 | −31.62 |
| $Manual_{cooked}$ | $L^*_{cooked}$ | 0.12 | 0.08 | −0.84 | -0.08 | −102.88 |
| $Manual_{cooked}$ | $a^*_{cooked}$ | 0.12 | 0.05 | 0.81 | 0.06 | 125.48 |
| $Manual_{cooked}$ | $b^*_{cooked}$ | 0.12 | 0.06 | −0.50 | −0.04 | −70.71 |
| $L^*_{raw}$ | $Manual_{uncooked}$ | 0.14 | 0.12 | −0.88 | −0.11 | −95.05 |
| $L^*_{raw}$ | $Manual_{cooked}$ | 0.14 | 0.12 | −0.81 | −0.10 | −87.49 |
| $L^*_{raw}$ | $L^*_{cooked}$ | 0.14 | 0.08 | 0.82 | 0.09 | 108.48 |
| $L^*_{raw}$ | $a^*_{cooked}$ | 0.14 | 0.05 | −0.81 | −0.07 | −135.54 |
| $L^*_{raw}$ | $b^*_{cooked}$ | 0.14 | 0.06 | 0.28 | 0.03 | 42.77 |
| $b^*_{raw}$ | $Manual_{uncooked}$ | 0.35 | 0.12 | −0.41 | −0.08 | −70.02 |
| $b^*_{raw}$ | $Manual_{cooked}$ | 0.35 | 0.12 | −0.54 | −0.11 | −92.22 |
| $b^*_{raw}$ | $L^*_{cooked}$ | 0.35 | 0.08 | 0.43 | 0.07 | 89.94 |
| $b^*_{raw}$ | $a^*_{cooked}$ | 0.35 | 0.05 | −0.53 | −0.07 | −140.22 |
| $b^*_{raw}$ | $b^*_{cooked}$ | 0.35 | 0.06 | 0.30 | 0.04 | 72.46 |
| $a^*_{raw}$ | $Manual_{uncooked}$ | 0.29 | 0.12 | 0.33 | 0.06 | 51.30 |
| $a^*_{raw}$ | $Manual_{cooked}$ | 0.29 | 0.12 | 0.36 | 0.07 | 55.96 |
| $a^*_{raw}$ | $L^*_{cooked}$ | 0.29 | 0.08 | −0.37 | −0.06 | −70.45 |
| $a^*_{raw}$ | $a^*_{cooked}$ | 0.29 | 0.05 | 0.10 | 0.01 | 24.08 |
| $a^*_{raw}$ | $b^*_{cooked}$ | 0.29 | 0.06 | −0.63 | −0.08 | −138.50 |
| Body weight | $Manual_{uncooked}$ | 0.27 | 0.12 | −0.28 | −0.05 | −42.00 |
| Body weight | $Manual_{cooked}$ | 0.27 | 0.12 | −0.30 | −0.05 | −45.00 |
| Body weight | $L^*_{uncooked}$ | 0.27 | 0.14 | 0.31 | 0.06 | 43.05 |
| Body weight | $a^*_{uncooked}$ | 0.27 | 0.29 | −0.21 | −0.06 | −20.26 |
| Body weight | $b^*_{uncooked}$ | 0.27 | 0.35 | 0.15 | 0.05 | 13.17 |
| Body weight | $L^*_{cooked}$ | 0.27 | 0.08 | 0.43 | 0.06 | 79.00 |
| Body weight | $a^*_{cooked}$ | 0.27 | 0.05 | −0.39 | −0.05 | −90.63 |
| Body weight | $b^*_{cooked}$ | 0.27 | 0.06 | 0.33 | 0.04 | 70.00 |
| $Manual_{uncooked}$ | Body weight | 0.12 | 0.27 | −0.28 | −0.05 | −18.67 |

$h^2$, heritabilty estimates; $r_g$, genetic correlation; IS, indirect selection efficiency as % of direct selection response.

precision of the data recording. For example, in our current study repeatability of scoring ($r$) ranged from 0.82 to 0.92. Instead, there are numerous reports where increased heritability estimates could be achieved by objective phenotyping of color traits using colorimetric instruments. Although, heritability estimates of manually-scored color phenotypes were low, it should be noted that, this manual approach may be very useful when instrumental methods are not available (Gjedrem, 2010), in addition it is fast and efficient and could be done on a processing line without additional handling or preparation needed for colorimetric measurements.

The effect of rearing environment on shrimp body coloration is well established. For example, Tume et al. (2009) reported that rearing of shrimp (*P. monodon*) for just 28 days, either in black or white tanks, had significant effect on body coloration. The black tank reared shrimp were more bright orange in color than the white tank color reared one. Moreover, Alam et al. (2022) found that shrimp (*P. monodon*) reared in pond containing mangrove leaf litters had significant darker body coloration. Our study also

revealed the variability of body coloration of *P. monodon* reared in different ponds, in particular cooked color, suggesting further study needs to be carried out to reveal the potential of genotype by environment (G×E) interaction for body color of this species.

From the six instrumental colorimetric measurements, heritability estimates ranged from 0.05 to 0.35. Similar to our findings, low to moderately high heritability value, ranging from 0.03 to 0.59, were also found in other shrimp species. Based on objective measurements of body color traits (e.g., lightness, yellowness and redness), Giang et al. (2019) reported heritability estimates ranging from 0.11–0.55, for body color traits of *Litopenaeus vannamei* reared in different environments. Nguyen et al. (2014) reported a heritability of 0.18 ± 0.05 and 0.08 ± 0.03 for uncooked and cooked color traits in banana shrimp (*Fenneropenaeus merguiensis*), respectively. In addition to shrimp, studies with other aquaculture species have shown that body color traits are heritable. Dufflocq et al. (2017), reported heritability estimates of 0.08 ± 0.02 and 0.04 ± 0.01 for flesh color traits, in two population of coho salmon (*Oncorhynchus kisutch*). For Atlantic salmon heritability estimates of 0.14 ± 0.03 (Tsai et al., 2015) and 0.07 ± 0.01 (Norris and Cunningham, 2004) were reported for color traits.

Of all the instrumental colorimetric measurements, heritability estimates were generally higher in the body color of the uncooked shrimp. For $a^*_{\text{uncooked}}$ and $b^*_{\text{uncooked}}$ values, the heritability estimates were 0.29 ± 0.08 and 0.35 ± 0.08, respectively, suggesting strong selection response will be obtained for these color traits. In particular, trait $a^*_{\text{uncooked}}$ can be a useful candidate trait for selection to improve redness color trait in shrimp, as previous studies have reported a linear relationship between carotenoid concentration and $a^*$ value in fish species, where carotenoid contents are the key contributor of bright coloration in crustaceans (Choubert, 1982; Skrede et al., 1990).

Surprisingly, heritability estimates were lower (ranged 0.05–0.08) for cooked color traits when body color phenotypes were measured instrumentally. This suggests that for genetic evaluation, instrumental measurements are more effective for uncooked body color traits than cooked traits. Given that both manually-scored and chronometrically-assessed body color traits identify significant additive genetic components, this suggests the higher potential of these traits for genetic improvement in *P. monodon*. Overall, our study confirms that the color phenotype has a substantial amount of genetic variation and is a good candidate trait for genetic improvement.

A key finding of our study was that the genetic correlation among the key body color traits were moderate to high, ranging from −0.88 to 0.81. Specifically, the genetic correlations between cooked and uncooked color traits were high (Table 3), suggesting that by selecting uncooked/live animals with desired color phenotypes, the cooked color of the animals can be improved genetically. This information is beneficial in shrimp breeding programs, as 1) it will eliminate the need for cooking the shrimp for phenotyping and thus will reduce overall associated costs in the breeding program; 2) will help breeders to preserve valuable genetic resource by not sacrificing them for cooked color phenotyping; and 3) measurements can be made at the same time as other important measurements are taken such as bodyweight and length due to moderately high genetic correlation between uncooked and cooked color of shrimp, it is predicted that selection for increased uncooked color (e.g., darker colored) will result in favorable change in cooked color (e.g., 77% at 1 SD selection intensity, Table 6) of shrimp.

Heritability of growth traits were moderately high in this study (e.g., 0.27 ± 0.07 for body weight and 0.32 ± 0.08 for body length traits), suggesting selection for these traits will lead to significant response in breeding programs. This finding corroborates with the previous findings, where heritability for growth traits were moderate to high, ranging from 0.23 to 0.69 in *P. monodon* (Hasan et al., 2020), indicating that these traits will be highly responsive during selection for genetic improvement. Of greater significance in this study is the impact of selection for growth on color traits. For the first time we show that color phenotypes were also moderate to highly correlated with growth traits in *P. monodon*. Similar finding was also reported by Giang et al. (2019) and Nguyen et al. (2014) for Pacific Whiteleg shrimp (*Litopenaeus vannamei*) and Banana shrimp (*Fenneropenaeus merguiensis*), respectively. Altogether these positive genetic correlations among color and growth traits suggests that a similar set of genes may be responsible for expression of these traits. Moreover, there might be physical linkage or pleiotropic effect and linkage disequilibrium among the underlying genetic mechanisms responsible for phenotypic expression of these traits (Falconer and Mackay, 1996). Similar findings of correlation with body color and growth traits have been reported in banana shrimp (Nguyen et al., 2014), salmonoids (Vieira et al., 2007) and in tilapia (Hamzah et al., 2016). The genetic correlations among body color and growth traits found from our study also indicate that there will be substantial correlated change when selection is applied on one trait or another. Selection for higher body weight will lead to a favorable increase of $L^*_{\text{cooked}}$ (108.48%) color of shrimp (Table 6). From a commercial perspective, this positive correlation between color and growth phenotypes suggests that color phenotype is highly suitable for shrimp aquaculture breeding program, since the growth combined with appealing coloration phenotype will increase overall profitability. Therefore, a selection index approach should be employed to simultaneously improve all the economically-important traits of the present population of shrimp, including growth, body shape and body color.

Studies with banana shrimp (Nguyen et al., 2014), salmon (Dufflocq et al., 2017) and tilapia (Hamzah et al., 2016), have also identified positive genetic correlations between body color

and growth traits. This suggests that similar genetic and metabolic pathways may be involved in regulating the expression of color phenotypes in shrimp. However, gene(s) that control color traits in shrimp remain unknown. In general pigmentation in crustacean species is determined by carotenoid astaxanthin which is mainly ingested by food sources and then converted and stabilized as protein crustacyanin in the tissue. A number of studies have demonstrated that levels of carotenoids in shrimp are positively correlated with growth and survival, suggesting individual shrimp with superior color characteristics are capable of converting carotenoids from the feed more efficiently and they grow better. Our findings, coupled with other studies with shrimp species, clearly demonstrates that body color traits are sufficiently heritable in *P. monodon*. This suggests that this trait can be genetically improved, implying that some individuals of the population studied possess superior ability to convert the carotenoid astaxanthin and also have superior growth compared to others. Of significance is that parameters of uncooked color can improve cooked color based on either subjective scores or objective colorimetric measures ($L^*_{\text{uncooked}}$ ,$a^*_{\text{uncooked}}$, $b^*_{\text{uncooked}}$). Finally, inclusion of cooked colour in the overall breeding objective for *P. monodon* will require clear economic benefits associated with the trait to be established through bio-economic modelling procedures as detailed by Marín-Riffo et al. (2021). Furthermore, all relevant genetic parameters and relative weights of all selection criteria, including uncooked colour and their relationship with the overall breeding objective will need to be established as detailed by Campos-Montes et al. (2017). However this is beyond the scope of the current study.

## Conclusion

Genetic parameter estimates for body color traits of *P. monodon* have been reported for the first time. The present study indicates that body color traits can respond effectively to selection. The high generic correlation between uncooked and cooked color scores indicates that, selection on dark colored (uncooked) shrimp will lead to enhanced intensity of cooked color. Moreover, positive genetic association among the growth and color traits indicates that, the selection for pigmentation and growth traits can be carried out simultaneously, without any unfavorable outcomes to these economically important traits. In summary, selective breeding can enhance growth and body color traits of shrimp simultaneously, thereby, helping to reduce the amount of food additives containing dietary astaxanthin. This will ultimately increase the overall product value and reduce feed costs.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

## Author contributions

All authors contributed to the manuscript. The study was conceived, designed and executed by MH, MK, HR, NW, DJ, and PT. The manuscript was conceived and prepared by MH, MK, PT, and HR. PT addressed statistical and grammatical components of the manuscript. The final manuscript was reviewed by all authors, and all approved its contents.

## Funding

This work was in part provided by the Australian Research Council–Industrial Transformation Research Hub (IH130200013) for Advanced Prawn Breeding. MH was supported by a Research Training Program Stipend (RTP), Francis Henry Loxton Supplementary Scholarship from the Sydney Institute of Agriculture, and The Paulette Isabel Jones PhD Completion Scholarship, at The University of Sydney.

## Acknowledgments

We acknowledge the support we received from all the funding bodies and collaborators with the acquisition of phenotypes, genotypes and data access, including all the members of ARC Research Hub for Advanced Prawn Breeding, James Cook University, Australia. The support from Seafarms Pty ltd. with the generation of the animals and access to farm resources is gratefully acknowledged.

## Conflict of interest

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

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their

affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1002346/full#supplementary-material

## References


Agresti, A. (2003). *Categorical data analysis*. John Wiley & Sons.

Alam, M. I., Yeasmin, S., Khatun, M. M., Rahman, M. M., Ahmed, M. U., Debrot, A. O., et al. (2022). Effect of mangrove leaf litter on shrimp (*Penaeus monodon*, Fabricius, 1798) growth and color. *Aquac. Rep.* 25, 101185. doi:10.1016/j.aqrep.2022.101185

Blay, C., Haffray, P., Bugeon, J., D'ambrosio, J., Dechamp, N., Collewet, G., et al. (2021). Genetic parameters and genome-wide association studies of quality traits characterised using imaging technologies in Rainbow trout, *Oncorhynchus mykiss*. *Front. Genet.* 12, 219. doi:10.3389/fgene.2021.639223

Boonyaratpalin, M., Thongrod, S., Supamattaya, K., Britton, G., and Schlipalius, L. (2001). Effects of β-carotene source,Dunaliella salina, and astaxanthin on pigmentation, growth, survival and health ofPenaeus monodon. *Aquac. Res.* 32, 182–190. doi:10.1046/j.1355-557x.2001.00039.x

Britton, G., Armitt, G. M., Lau, S. Y., Patel, A. K., and Shone, C. C. (1982). "Carotenoproteins," in Carotenoid Chemistry and biochemistry*)*, 237–251.

Butler, D., Cullis, B. R., Gilmour, A., and Gogel, B. (2017). *ASReml-R reference manual Version 4*. Hemel Hempstead, United Kingdom: VSN International Ltd.

Calo, L., McDowell, R., VanVleck, L. D., and Miller, P. (1973). Genetic aspects of beef production among holstein-friesians pedigree selected for milk Production1. *J. Animal Sci.* 37 (3), 676–682. doi:10.2527/jas1973.373676x

Campos-Montes, G. R., Montaldo, H. H., Armenta-Córdova, M., Martínez-Ortega, A., Caballero-Zamora, A., and Castillo-Juárez, H. (2017). Incorporation of tail weight and tail percentage at harvest size in selection programs for the Pacific white shrimp Penaeus ( Litopenaeus ) vannamei. *Aquaculture* 468, 293–296. doi:10.1016/j.aquaculture.2016.10.034

Choubert, G. (1982). Method for colour assessment of canthaxanthin pigmented rainbow trout (*Salmo gairdneri* Rich.). *Sci. Aliments* 2 (4), 451–463.

Christiansen, R., Struksnæs, G., Estermann, R., and Torrissen, O. (1995). Assessment of flesh colour in Atlantic salmon, *Salmo salar* L. *Aquac. Res.* 26 (5), 311–321. doi:10.1111/j.1365-2109.1995.tb00919.x

de Carvalho, C. C., and Caramujo, M. J. (2017). Carotenoids in aquatic ecosystems and aquaculture: A colorful business with implications for human health. *Front. Mar. Sci.* 4, 93. doi:10.3389/fmars.2017.00093

Dufflocq, P., Lhorente, J. P., Bangera, R., Neira, R., Newman, S., and Yánez, J. M. (2017). Correlated response of flesh color to selection for harvest weight in coho salmon (*Oncorhynchus kisutch*). *Aquaculture* 472, 38–43. doi:10.1016/j.aquaculture.2016.08.037

Falconer, D., and Mackay, T. (1996). *Introduction to quantitative genetics*. Essex. UK: Longman Group.

FAO (2020). *The state of world Fisheries and aquaculture: Sustainability in action*. Rome, Italy: Food and Agriculture Organization of the United Nations. Retrived from https://www.fao.org/3/ca9229en/ca9229en.pdf.

Foote, A., Simma, D., Khatkar, M., Raadsma, H., Guppy, J., Coman, G., et al. (2019). Considerations for maintaining family diversity in commercially mass-spawned penaeid shrimp: A case study on *Penaeus monodon*. *Front. Genet.* 10, 1127. doi:10.3389/fgene.2019.01127

Giang, C. T., Knibb, W., Muu, N. H., Ninh, N. H., and Nguyen, N. H. (2019). Prospects for genetic improvement in objective measurements of body colour in Pacific Whiteleg shrimp (*Litopenaeus vannamei*). *Jmse* 7 (12), 460. doi:10.3390/jmse7120460

Gjedrem, T. (2010). The first family-based breeding program in aquaculture. *Rev. Aquac.* 2 (1), 2–15. doi:10.1111/j.1753-5131.2010.01011.x

Gómez-Polo, C., Muńoz, M. P., Luengo, M. C. L., Vicente, P., Galindo, P., and Casado, A. M. M. (2016). Comparison of the CIELab and CIEDE2000 color difference formulas. *J. Prosthet. Dent.* 115 (1), 65–70.

Guppy, J. L., Jones, D. B., Kjeldsen, S. R., Le Port, A., Khatkar, M. S., Wade, N. M., et al. (2020). Development and validation of a RAD-Seq target-capture based genotyping assay for routine application in advanced black tiger shrimp (*Penaeus monodon*) breeding programs. *BMC Genomics* 21 (1), 541. doi:10.1186/s12864-020-06960-w

Hamzah, A., Nguyen, N. H., Mekkawy, W., Ponzoni, R. W., Khaw, H. L., Yee, H. Y., et al. (2016). Flesh characteristics: Estimation of genetic parameters and correlated responses to selection for growth rate in the GIFT strain. *Aquac. Res.* 47 (7), 2139–2149. doi:10.1111/are.12667

Hasan, M. M. (2022). *Genetic investigation of body morphological traits in* Penaeus monodon *and their relationship to shrimp production*. Australia: Doctoral dissertation, The University of Sydney.

Hasan, M. M., Tulloch, R. L., Thomson, P. C., Raadsma, H. W., and Khatkar, M. S. (2020). Meta-analysis of genetic parameters of production traits in cultured shrimp species. *Fish. Fish.* 21 (6), 1150–1174. doi:10.1111/faf.12495

Jones, O. R., and Wang, J. (2010). Colony: A program for parentage and sibship inference from multilocus genotype data. *Mol. Ecol. Resour.* 10 (3), 551–555. doi:10.1111/j.1755-0998.2009.02787.x

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16 (5), 1099–1106. doi:10.1111/j.1365-294x.2007.03089.x

Kause, A., Ritola, O., Paananen, T., Mäntysaari, E., and Eskelinen, U. (2002). Coupling body weight and its composition: A quantitative genetic analysis in rainbow trout. *Aquaculture* 211 (1-4), 65–79. doi:10.1016/s0044-8486(01)00884-5

Ly, B. C. K., Dyer, E. B., Feig, J. L., Chien, A. L., and Del Bino, S. (2020). Research techniques made simple: Cutaneous colorimetry: A reliable technique for objective skin color measurement. *J. Investigative Dermatology* 140 (1), 3–12. e11. doi:10.1016/j.jid.2019.11.003

Marín-Riffo, M. C., Raadsma, H. W., Jerry, D. R., Coman, G. J., and Khatkar, M. S. (2021). Bioeconomic modelling of hatchery, grow-out and combined business of Australian black tiger shrimp *Penaeus monodon* farming. *Rev. Aquac.* 13 (3), 1695–1708.

Mehar, M., Mekkawy, W., McDougall, C., and Benzie, J. A. (2020). Fish trait preferences: A review of existing knowledge and implications for breeding programmes. *Rev. Aquac.* 12 (3), 1273–1296.

Menasveta, P., Worawattanamateekul, W., Latscha, T., and Clark, J. (1993). Correction of black tiger prawn (*Penaeus monodon* Fabricius) coloration by astaxanthin. *Aquac. Eng.* 12 (4), 203–213. doi:10.1016/0144-8609(93)90012-z

Nguyen, N. H., Quinn, J., Powell, D., Elizur, A., Thoa, N. P., Nocillado, J., et al. (2014). Heritability for body colour and its genetic association with morphometric traits in Banana shrimp (*Fenneropenaeus merguiensis*). *BMC Genet.* 15 (1), 1–12. doi:10.1186/s12863-014-0132-5

Niu, J., Wen, H., Li, C.-H., Liu, Y.-J., Tian, L.-X., Chen, X., et al. (2014). Comparison effect of dietary astaxanthin and β-carotene in the presence and absence of cholesterol supplementation on growth performance, antioxidant capacity and gene expression of *Penaeus monodon* under normoxia and hypoxia condition. *Aquaculture* 422-423, 8–17. doi:10.1016/j.aquaculture.2013.11.013

Noble, T. H., Coman, G. J., Wade, N. M., Thomson, P. C., Raadsma, H. W., Khatkar, M. S., et al. (2020). Genetic parameters for tolerance to gill-associated virus under challenge-test conditions in the black tiger shrimp (*Penaeus monodon*). *Aquaculture* 516, 734428. doi:10.1016/j.aquaculture.2019.734428

Norris, A., and Cunningham, E. (2004). Estimates of phenotypic and genetic parameters for flesh colour traits in farmed Atlantic salmon based on multiple trait animal model. *Livest. Prod. Sci.* 89 (2-3), 209–222. doi:10.1016/j.livprodsci.2004.02.010

Paibulkichakul, C., Piyatiratitivorakul, S., Sorgeloos, P., and Menasveta, P. (2008). Improved maturation of pond-reared, black tiger shrimp (*Penaeus monodon*) using fish oil and astaxanthin feed supplements. *Aquaculture* 282 (1-4), 83–89. doi:10.1016/j.aquaculture.2008.06.006

Peshanoff, L., and Jaensch, J. (2009). *Consumer research: Australian Prawn farmers association*. Adelaide, Australia: University of South Australia.

R Core Team (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Robertson, A. (1977). The CIE 1976 color-difference formulae. *Color Res. Appl.* 2 (1), 7–11. doi:10.1002/j.1520-6378.1977.tb00104.x

Rye, M., and Gjerde, B. (1996). Phenotypic and genetic parameters of body composition traits and flesh colour in Atlantic salmon, *Salmo salar* L. *Aquac. Res.* 27, 121–133.

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., et al. (2011). Diversity arrays technology (DArT) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proc.* 5 (7), 1–2. doi:10.1186/1753-6561-5-s7-p54

Skrede, G., Risvik, E., Huber, M., Enersen, G., and Blümlein, L. (1990). Developing a color card for raw flesh of astaxanthin-fed salmon. *J Food Sci.* 55 (2), 361–363. doi:10.1111/j.1365-2621.1990.tb06763.x

Supamattaya, K., Kiriratnikom, S., Boonyaratpalin, M., and Borowitzka, L. (2005). Effect of a *Dunaliella* extract on growth performance, health condition, immune response and disease resistance in black tiger shrimp (*Penaeus monodon*). *Aquaculture* 248 (1-4), 207–216. doi:10.1016/j.aquaculture.2005.04.014

Tsai, H. Y., Hamilton, A., Guy, D. R., Tinch, A. E., Bishop, S. C., and Houston, R. D. (2015). The genetic architecture of growth and fillet traits in farmed Atlantic salmon (*Salmo salar*). *BMC Genet.* 16 (1), 51–11. doi:10.1186/s12863-015-0215-y

Tume, R., Sikes, A., Tabrett, S., and Smith, D. (2009). Effect of background colour on the distribution of astaxanthin in black tiger prawn (*Penaeus monodon*): Effective method for improvement of cooked colour. *Aquaculture* 296 (1-2), 129–135. doi:10.1016/j.aquaculture.2009.08.006

Vieira, V. L., Norris, A., and Johnston, I. A. (2007). Heritability of fibre number and size parameters and their genetic relationship to flesh quality traits in Atlantic salmon (*Salmo salar* L.). *Aquaculture* 272, S100–S109. doi:10.1016/j.aquaculture.2007.08.028

Wade, N., and Glencross, B. (2011). "Optimising External Colour in Farmed Crustaceans, using *Penaeus monodon* as a model species," in *Seafood CRC project*. Retrived from https://www.seafoodcrc.com/components/com_virtuemart/attachments/2011731.pdf.

Wade, N. M., Anderson, M., Sellars, M. J., Tume, R. K., Preston, N. P., and Glencross, B. D. (2012). Mechanisms of colour adaptation in the prawn *Penaeus monodon*. *J. Exp. Biol.* 215 (2), 343–350. doi:10.1242/jeb.064592

Wade, N. M., Gabaudan, J., and Glencross, B. D. (2017). A review of carotenoid utilisation and function in crustacean aquaculture. *Rev. Aquacult* 9 (2), 141–156. doi:10.1111/raq.12109

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 13 October 2022
DOI 10.3389/fgene.2022.1018568



OPEN ACCESS

EDITED BY
Yuzine Esa,
Putra Malaysia University, Malaysia

REVIEWED BY
Florence Phocas,
INRAE Centre Jouy-en-Josas, France
Dania Aziz,
Universiti Putra Malaysia, Malaysia

*CORRESPONDENCE
Shengjie Ren,
shengjieren@hotmail.com

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

RECEIVED 13 August 2022
ACCEPTED 26 September 2022
PUBLISHED 13 October 2022

CITATION
Ren S, Mather PB, Tang B and Hurwood DA (2022), Insight into selective breeding for robustness based on field survival records: New genetic evaluation of survival traits in pacific white shrimp (Penaeus vannamei) breeding line.
*Front. Genet.* 13:1018568.
doi: 10.3389/fgene.2022.1018568



# Insight into selective breeding for robustness based on field survival records: New genetic evaluation of survival traits in pacific white shrimp (*Penaeus vannamei*) breeding line

Shengjie Ren[1]*, Peter B. Mather[1], Binguo Tang[2] and David A. Hurwood[1]

[1]Faculty of Science, Queensland University of Technology, Brisbane, QLD, Australia, [2]Beijing Shuishiji Biotechnology Co., Ltd., Beijing, China

Survival can be considered a relatively 'old' trait in animal breeding, yet commonly neglected in aquaculture breeding because of the simple binary records and generally low heritability estimates. Developing routine genetic evaluation systems for survival traits however, will be important for breeding robust strains based on valuable field survival data. In the current study, linear multivariate animal model (LMA) was used for the genetic analysis of survival records from 2-year classes (BL2019 and BL2020) of pacific white shrimp (*Penaeus vannamei*) breeding lines with data collection of 52, 248 individuals from 481 fullsib families. During grow-out test period, 10 days intervals of survival data were considered as separate traits. Two survival definitions, binary survivability (S) and continuous survival in days (SL), were used for the genetic analysis of survival records to investigate; 1) whether adding more survival time information could improve estimation of genetic parameters; 2) the trajectory of survival heritability across time, and 3) patterns of genetic correlations of survival traits across time. Levels of heritability estimates for both S and SL were low (0.005–0.076), while heritability for survival day number was found to be similar with that of binary records at each observation time and were highly genetically correlated ($r_g$ > 0.8). Heritability estimates of body weight (BW) for BL2019 and BL2020 were 0.486 and 0.373, respectively. Trajectories of survival heritability showed a gradual increase across the grow-out test period but slowed or reached a plateau during the later grow-out test period. Genetic correlations among survival traits in the grow-out tests were moderate to high, and the closer the times were between estimates, the higher were their genetic correlations. In contrast, genetic correlations between both survival traits and body weight were low but positive. Here we provide the first report on the trajectory of heritability estimates for survival traits across grow-out stage in aquaculture. Results will be useful for developing robust improved pacific white shrimp culture strains in selective breeding programs based on field survival data.

KEYWORDS

*Penaeus vannamei*, survival traits, genetic parameters, heritability, genetic evaluation, selective breeding

## Introduction

Aquaculture is playing an increasingly important role in world food security, development of economic sustainability, and to provide practical solutions for addressing ecosystem services issues (Bernatchez et al., 2017; Houston et al., 2020; Naylor et al., 2021). Sustainability of aquaculture production in global food systems however, has become vulnerable due to the rapid expansion of aquaculture industry, outbreak of diseases/pathogens, water environmental pollution, and in particular, from the impacts of climate changes (Troell et al., 2014; Reid et al., 2019). Consequently, there has seen an increasing demand for better management and breeding of robust culture lines (Friggens et al., 2017). Genetic improvement *via* selective breeding is widely acknowledged as an efficient tool to improve economically important traits including: feed conversion ratios, biomass production, and overall survival rates of domesticated aquatic animals (Gjedrem et al., 2012; Hung et al., 2013; Nguyen, 2016; Gjedrem and Rye, 2018). A wide range of projects have confirmed selective breeding to be an efficient strategy for enhancing overall survival rate and disease resistance performance in farmed aquatic species, and thereby contributing to the development of aquaculture in a sustainable way (Yáñez et al., 2014; Houston, 2017).

In general terms, two selection approaches have been commonly used to develop robustness in aquaculture genetic breeding programs. One approach is to improve specific disease resistance *via* controlled challenge tests in target breeding lines. For this approach, tested families are artificially infected with a specific pathogen in a controlled environment condition *via* intra-peritoneal injection, immersion or cohabitation. Following this, additive genetic variance among families for phenotypic resistance against the target pathogen can be identified and used in future breeding plans (Ødegård et al., 2011; Yáñez et al., 2014). Currently, disease resistance strains have been developed successfully *via* this approach for a number of aquatic species, including farmed salmonid species (Correa et al., 2015; Vallejo et al., 2017; Barría et al., 2019), Pacific oyster (Gutierrez et al., 2018), and European sea bass (Palaiokostas et al., 2018). A second approach is to use selection based on survival data records in the field and this can provide another important data source for selecting robustness by improving overall individual survival rate (Gjedrem and Rye, 2018) and for developing specific disease resistance strains (Barría et al., 2020; Fraslin et al., 2022). This approach allows direct collection of data under real commercial farm conditions, and thereby avoids potential for genotype-by-environment (G-by-E) problems as seen in controlled challenge experiments between challenge test environments and production conditions on farm. Survival data records however, are often neglected in aquaculture breeding programs because they are based on simple binary records (0 or 1) and generally show low heritability.

Genetic analyses of survival phenotypic data in aquaculture breeding programs are commonly treated as a binary trait with 'alive vs. dead' reported for individuals at a specific observation time point scored as 1 and 0, respectively. Under this scenario, individuals that died early or later during the grow-out period would be given the same score, which means that useful information about relative survival time and/or lifespan is lost and therefore has not been used in the genetic analysis (Ødegård et al., 2006). Moreover, binary record variables of survival are commonly non-normally distributed, an issue that may compromise estimations of genetic components using a linear mixed animal model. To address this problem, survival records scored as continuous traits of survival time have been implemented successfully for genetic evaluation of survival data in animal breeding programs (Ducrocq and Casella, 1996; Van Pelt et al., 2015; Heise et al., 2016). In this way, survival phenotype can be recorded as normally distributed continuous traits and information about individuals with different survival time/lifespans can be captured effectively in the analysis. To date, genetic analyses of survival data scored as continuous traits have only been reported on farmed aquatic species for controlled challenge tests (Suebsong et al., 2019; Joshi et al., 2021a) while survival data from commercial breeding lines has rarely been investigated.

Different genetic evaluation models are available for genetic analyses of survival traits in animal breeding (Forabosco et al., 2009). A proportional hazard model (PHM) is a common model used for genetic evaluation of survival traits, and can be applied with a free open access software package in *Survival Kit* (Ducrocq and Sölkner, 1994; Ducrocq et al., 2010). PHM can handle large survival data sets rapidly and easily, manage survival data sets with skewed distributions, and is considered to provide accurate estimates (Sewalem et al., 2010; Zavadilová et al., 2011). In genetic evaluations however, it is difficult to process genetic correlations with other continuous traits at the same time (Tarrés et al., 2006). Alternatively, linear models and threshold models have been widely used for genetic analyses of survival data in aquaculture genetics. Compared with linear models, threshold models applying a logit link function are feasible for dealing with non-normally distributed binary survival data. However, they generally require more computational time and cannot estimate genetic correlations with other continuous traits simultaneously. In practice, analyses using either threshold models or PHM take almost five to ten times more computational time than do applying linear models to address almost the same tasks (Boettcher et al.,

1999). Moreover, estimations of true breeding values (EBVs) have shown very similar correlations between linear models and threshold models (Veerkamp et al., 2002). While different linear models including random regression models (RRMs) have been used successfully for national genetic evaluations of survival data in dairy cattle (Sasaki et al., 2015; Van Pelt et al., 2015, 2016; Heise et al., 2016, 2018), there are currently no standardized model choices for routine genetic evaluation of survival data in aquaculture breeding programs.

Another important consideration for genetic analysis of survival traits in aquaculture species is changes of time. Survival traits are dynamic quantitative traits, which can change spatially and temporally due to multiple interactions between animal and environmental constraints. Therefore, understanding the trajectory of time for survival traits can be very useful for making critical decisions in a breeding plan (Schaeffer, 2004). Studies on trajectory of time for important economic traits in aquaculture genetics are still rare (Vehviläinen et al., 2010), and the few reported cases have mostly focused on growth traits (Turra et al., 2012; He et al., 2017; Schlicht et al., 2018), while to date, there has been no reports on trajectory of survival traits across grow-out testing stage.

Pacific white shrimp (*Penaeus vannamei*) has become the most widely farmed prawn species across the world. Annual global production reached ~4.4 million tons with a commercial value of 26.7 billion USD in 2020, which ranked as the most important traded food commodity across the aquaculture sector (Kumar and Engle, 2016; FAO, 2020). Sustainability of prawn farming has been affected however, by emergence of several diseases that show high mortality rates (Robinson et al., 2022). Developing robust culture strains of pacific shrimp *via* selective breeding will play a crucial role for improving the economic profitability and animal welfare of this major aquatic farmed species. The main objective here was therefore to develop routine genetic evaluation of survival records in a genetic improvement program for pacific shrimp in China. Specifically, we evaluated: 1) genetic parameters from two different survival definitions for the binary traits of survivability and the continuous trait of survival time; 2) the trajectory of heritability for survival traits across the grow-out test period; and 3) genetic correlation patterns for survival traits across grow-out stage and their correlation with growth. Results from the current program will provide insight into selection breeding of robust culture strains based on field survival records in aquaculture species.

# Materials and methods

## Study population

The study population constituted the breeding nucleus from a pacific white shrimp stock improvement program in Hainan Island (China), with the selection target being for local farm environments in China based on a family selection approach. Foundation populations were produced in 2015 sourced from 12 hatchery lines in China representing four genetic populations as evidenced from a population structure analysis (Ren et al., 2018). In 2015, 98 full families were produced (Ren et al., 2020a). Following this, for each breeding cycle, 209–250 families were produced over a 1-week period. Family pedigree management used physical visible implant elastomer (VIE) tags, while in parallel a parentage assignment panel (Ren et al., 2022) was developed in 2019 to meet the demands of more large family numbers in the selective breeding program. The mating system for this program used a nested mating design *via* one single male with two females. The ratio of dam/sire was maintained at ~ 1.7 with the aim to generate better genetic tier for EBVs estimation. Grow-out management conditions of the breeding line have been reported in earlier studies (Ren et al., 2020b, Ren et al., 2020c).

## Data records

In the current study, genetic analysis data were sourced from two different year classes 2019 (BL 2019) and 2020 (BL 2020) in the breeding nucleus line. The BL2019 line, consisted of 243 full-sibs families generated from 157 sires and 243 dams over a 1-week period, while the BL2020 line consisted of 238 full-sibs families produced by 143 sires and 238 dams (Table 1). The mean number of shrimp per family for the grow-out test was 102.8 for BL2019 and 114.7 for BL2020, with a total number of 52, 248 individuals recorded for data collection (Table 1).

Dead individuals from the breeding lines were collected three times per day across the grow-out period and pedigree information of mortalities was recorded from visible implant elastomer (VIE) tags. At the end of the grow-out test stage, all remaining harvested individuals in the test system were scored for body weight, pedigree of family ID, and gender.

## Trait definition

Data collected on survival from each 10 days interval were considered as separate traits over the grow-out period. Survival phenotype was recorded using two definitions in the genetic analysis. First, survival phenotype at the end of each 10 day period was coded as a binary trait of survivability, with 0 for dead individuals and 1 for live individuals. Secondly, survival data for each 10 day period was considered to be a continuous trait of survival time indicating how many days individuals survived over the test period. For example, live individuals in each breeding line at surviving to 100 days were coded as 100d and dead individuals at 86 days or 56 days were coded as 86d or 56d, respectively. Therefore, survival data records from BL2019 breeding line across the 117 days grow-out period were coded as

TABLE 1 Data structure of pacific white shrimp breeding nucleus lines.

| Year | No. Shrimp | Shrimp/family | No. Family | Sires | Dams |
|---|---|---|---|---|---|
| 2019 | 24, 980 | 102.8 | 243 | 157 | 243 |
| 2020 | 27, 304 | 114.7 | 238 | 143 | 238 |
| Total/mean | 52, 284 | 108.7 | 481 | 300 | 481 |

12 survivability binary traits (S1, S2, . . ., S11, S12) and 12 continuous traits of survival (SL1, SL2, . . ., SL11, SL12) for each observed time window of 10 days. Similarly, the 98 days records of survival data from BL2020 considered 10 survival binary traits (S1, S2, . . ., S9, S10) and 10 continuous traits of survival days (SL1, SL2, . . ., SL9, SL10), respectively. At the conclusion of the grow-out tests, phenotype of body weight for each shrimp was collected as a continuous trait (BW), while data from the final population census of live shrimps were recorded as a binary survival trait (SUR).

## Statistical analysis

### Survival probability estimates and daily mortality

Changes in survival probability of shrimp in breeding lines across the experimental grow-out period were analysed using the Kaplan-Meier estimator (Kaplan and Meier, 1958). Shrimp that survived at each observed time (days) were treated as censored. The "Survminer" function (Kassambara et al., 2017) in the **R** package (R Core Team, 2019) was used to estimate Kaplan-Meier survival curves by;

$$\hat{S}(t) = \frac{\hat{\Pi}}{t_i \le t}\left(1 - \frac{d_i}{n_i}\right)$$

where, $n_i$ is the number of alive shrimp at risk at observed time $t_i$, and $d_i$ is the number of deaths at the observed time. In addition, daily mortality in the breeding lines was also plotted for the general husbandry management assessment purposes.

## Genetic analyses

### Genetic parameters

WOMBAT software (Meyer, 2007) was used to fit the following linear multivariate animal model (LMA) for the genetic analysis:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{e} \quad (1)$$

where, $\boldsymbol{y}$ is a vector of phenotype for the traits defined here (S1-S12, SL1-SL12, SUR, and BW); $\boldsymbol{\beta}$ is the vector of fixed effects including sex, tanks and family batches; $\boldsymbol{\alpha}$ is the vector of random additive genetic effects; $\boldsymbol{e}$ is the vector of random residual errors; and, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known incidence matrices relating observations to the fixed and random effects mentioned above. Both vector of $\boldsymbol{\alpha}$ and $\boldsymbol{e}$ are assumed to be multivariate normal distribution with mean zero and variances as:

$$\mathrm{Var}\begin{bmatrix}\alpha\\ e\end{bmatrix} = \begin{bmatrix}\mathbf{A}\boldsymbol{\sigma}_{\boldsymbol{\alpha}}^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\boldsymbol{\sigma}_{\alpha}^2\end{bmatrix},$$

here, $\boldsymbol{\sigma_\alpha}^2$ and $\boldsymbol{\sigma_e}^2$ are the random additive variances and error variances, respectively. **A** is the numerator relationship matrix based on pedigree information, and **I** represents an identity matrix. Total variance ($\sigma_p{}^2$) was calculated as the sum of random additive genetic variance ($\sigma_\alpha{}^2$) and random residual components ($\sigma_e{}^2$). Heritability ($h^2$) was calculated as the ratio of the random genetic variance to the total phenotype variance, $h^2 = \sigma_\alpha^2/\sigma_p^2$. Phenotypic correlation ($r_p$) between two traits was calculated as: $\gamma_p = \frac{\sigma_{\alpha 1,2}+\sigma_{e1,2}}{\sqrt{(\sigma_{\alpha 1}^2+\sigma_{e1}^2)+(\sigma_{\alpha 2}^2+\sigma_{e2}^2)}}$; and genetic correlation ($r_g$) between two traits was calculated as:$\gamma_g = \frac{\sigma_{\alpha 1,2}}{\sqrt{\sigma_{\alpha 1}^2\,\sigma_{\alpha 2}^2}}$. The values of data matrix for $r_p$ and $r_g$ were displayed as heat maps in **R** package (R Core Team, 2019). From the output of LMA genetic analyses, $r_p$ and $r_g$ between phenotype traits were aligned using **R** package in the 'Matrix' package and after, function of 'pheatmap' in R was used to plot the graphic representations of the correlation matrix data.

## Correlations of binary traits and survival days

Bivariate analyses were performed, in which the two definition types of survival phenotype were modeled simultaneously. Equations used in the bivariate animal model can be defined as follows:

$$\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}\boldsymbol{X} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{X}\end{bmatrix}\begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{Z} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{Z}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\end{bmatrix} + \begin{bmatrix}\boldsymbol{e}_1\\ \boldsymbol{e}_2\end{bmatrix} \quad (2)$$

where, the symbols represent the same vectors as described in the multivariate analysis of Model 1; the subscripts 1 and 2 are two different records for survival data of binary traits (S1-S12) and survival days (SL1-SL12) at each observed time window, respectively.

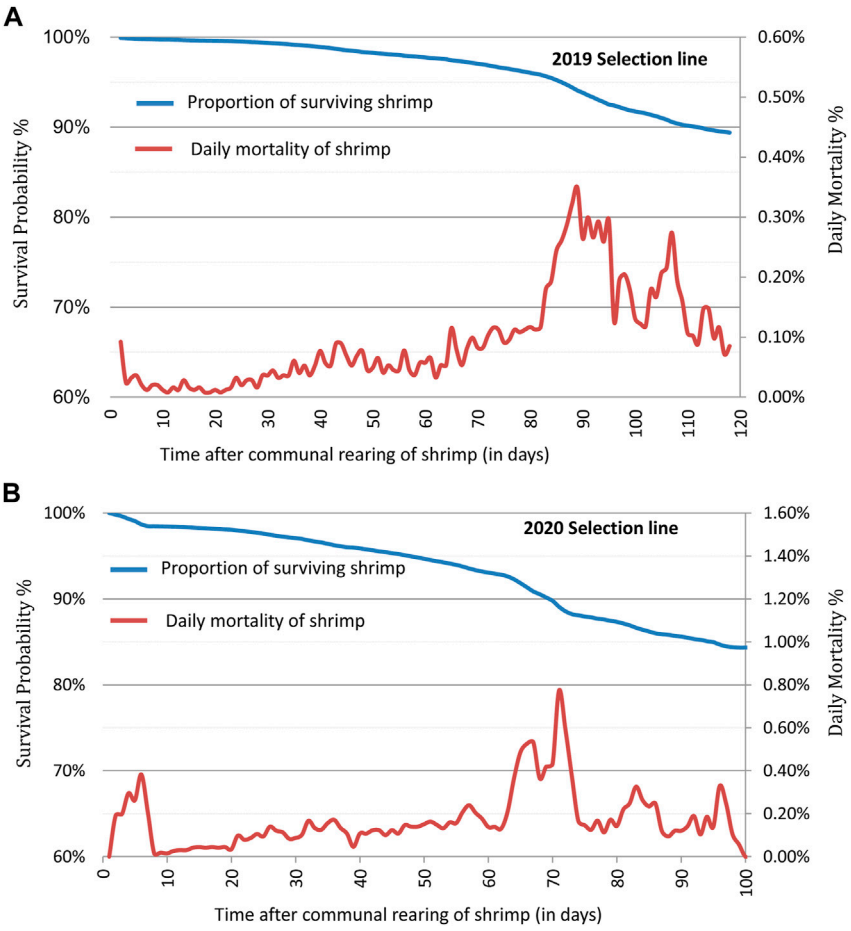



FIGURE 1
Kaplan-Meier survival curves and daily mortality changes for the shrimp breeding lines: **(A)** 2019 (BL 2019) of 243 fullsib families and **(B)** 2020 (BL 2020) of 238 fullsib families.

## Common environmental effect

Two linear univariate animal models were developed for the significance test of the full-sib family effect ($c^2$) in the genetic analyses. Equations of univariate animal models can be written as follows:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{Wf} + \boldsymbol{e} \quad (3)$$

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\alpha} + \boldsymbol{e} \quad (4)$$

where, $\boldsymbol{f}$ is the vector of random full-sib family effects and $\mathbf{W}$ is the corresponding design matrices; other symbols represent the same vectors as described in Model 1. Compared with Model 3, there was no random full-sib family effect in Model 4. Likely statistical significance for the random full-sib family effects were analysed using a likelihood ratio approach. After running the two above univariate Models, significance for the random full-sib family effects were compared to the final log-likelihood (Maximum log L) using a Chi-square ($\chi^2$) test.

# Results

## Patterns of survival

Kaplan–Meier survival curves showed changes in survival probability across the grow-out test periods for each year breeding line over the 2 year test period (Figure 1). At the end of the grow-out test period, survival probability for BL2019 was 89.4%, while for BL2020 it was 84.4%. These survival estimators however, presented some bias to the final population census (80.4% and 64.2%) at the end of grow-out test. Differences between the bias estimates indicate a proportion of shrimp mortality events apparently were not detected over the test periods. There were also significant differences ($p < 0.01$) in mortality among families for the two breeding lines investigated here. Means of daily mortality records for BL2019 and BL2020 were 0.1% and 0.17%, respectively (Figure 1). Overall, daily mortalities fluctuated at 0.1%–0.25% for breeding lines

TABLE 2 Descriptive statistics of phenotype traits for the breeding lines of BL2019 and BL 2020.

| Trait | *BL2019* | | | | *BL2020* | | | |
|---|---|---|---|---|---|---|---|---|
| | ***N*** | **Min** | **Max** | **Mean** | ***N*** | **Min** | **Max** | **Mean** |
| S1 | 24,980 | 0.000 | 1.000 | 0.998 | 27,304 | 0.000 | 1.000 | 0.984 |
| S2 | 24,980 | 0.000 | 1.000 | 0.996 | 27,304 | 0.000 | 1.000 | 0.981 |
| S3 | 24,980 | 0.000 | 1.000 | 0.994 | 27,304 | 0.000 | 1.000 | 0.972 |
| S4 | 24,980 | 0.000 | 1.000 | 0.990 | 27,304 | 0.000 | 1.000 | 0.959 |
| S5 | 24,980 | 0.000 | 1.000 | 0.984 | 27,304 | 0.000 | 1.000 | 0.947 |
| S6 | 24,980 | 0.000 | 1.000 | 0.979 | 27,304 | 0.000 | 1.000 | 0.932 |
| S7 | 24,980 | 0.000 | 1.000 | 0.973 | 27,304 | 0.000 | 1.000 | 0.899 |
| S8 | 24,980 | 0.000 | 1.000 | 0.963 | 27,304 | 0.000 | 1.000 | 0.876 |
| S9 | 24,980 | 0.000 | 1.000 | 0.942 | 27,304 | 0.000 | 1.000 | 0.865 |
| S10 | 24,980 | 0.000 | 1.000 | 0.922 | 27,304 | 0.000 | 1.000 | 0.859 |
| S11 | 24,980 | 0.000 | 1.000 | 0.910 | *NA* | *NA* | *NA* | *NA* |
| S12 | 24,980 | 0.000 | 1.000 | 0.903 | *NA* | *NA* | *NA* | *NA* |
| SL1 | 24,980 | 1.000 | 10.000 | 9.986 | 27,304 | 1.000 | 10.000 | 9.904 |
| SL2 | 24,980 | 1.000 | 20.000 | 19.957 | 27,304 | 1.000 | 20.000 | 19.730 |
| SL3 | 24,980 | 1.000 | 30.000 | 29.909 | 27,304 | 1.000 | 30.000 | 29.493 |
| SL4 | 24,980 | 1.000 | 40.000 | 39.833 | 27,304 | 1.000 | 40.000 | 39.134 |
| SL5 | 24,980 | 1.000 | 50.000 | 49.704 | 27,304 | 1.000 | 50.000 | 48.657 |
| SL6 | 24,980 | 1.000 | 60.000 | 59.521 | 27,304 | 1.000 | 60.000 | 58.045 |
| SL7 | 24,980 | 1.000 | 70.000 | 69.288 | 27,304 | 1.000 | 70.000 | 67.215 |
| SL8 | 24,980 | 1.000 | 80.000 | 78.978 | 27,304 | 1.000 | 80.000 | 76.058 |
| SL9 | 24,980 | 1.000 | 90.000 | 88.535 | 27,304 | 1.000 | 90.000 | 84.753 |
| SL10 | 24,980 | 1.000 | 100.000 | 97.862 | 27,304 | 1.000 | 98.000 | 93.352 |
| SL11 | 24,980 | 1.000 | 110.000 | 107.024 | *NA* | *NA* | *NA* | *NA* |
| SL12 | 24,980 | 1.000 | 118.000 | 116.089 | *NA* | *NA* | *NA* | *NA* |
| SUR | 24,980 | 0.000 | 1.000 | 0.755 | 27,304 | 0.000 | 1.000 | 0.585 |
| BW | 20,078 | 2.300 | 57.000 | 21.394 | 17,529 | 1.900 | 41.900 | 17.946 |

*N*, the number of shrimp; S1-S12, survivability of binary traits for each observed time window of 10 days period; SL1-SL12, continuous traits of survival days for each observed time window of 10 days period; SUR, binary trait of survival for the final population census; BW, body weight of shrimp.

across the 2 years. Small peaks in mortality were first observed at the beginning of each test in the first week, which in part, can be explained as due to handling stress during VIE tagging. Most daily mortality peaks were evident at later stages of the grow-out test periods, where the highest daily mortality at 0.38% peaked at ~ 90 days for BL 2019, while the highest daily mortality for BL2020 peaked for 0.78% at 70 days. The time period of the highest daily mortality events coincided with the coldest winter season period at the hatchery location in Hainan.

## Descriptive statistics of genetic analyses

For both year breeding lines, all families had records for survival data. Basic descriptive statistics for each trait are presented in Table 2. The coefficient of variation (CV) for survival traits gradually increased for survival data record (SL) in both year lines. CV for survival days ranged from 3.32% to 12.49% for BL 2019, while similar patterns for CV of SL were also observed for the BL 2020 line. Means of survival trait (S) at the end of the grow-out test period were higher than that of SUR, indicating some mortality events had not been observed.

## Genetic parameters

Estimates of variance components and heritability of survivability records for binary traits (S), survival days (SL), final population census of survival (SUR), and body weight (BW) are presented in Table 3. Levels of heritability estimates for S and SL were both low and ranged from 0.005 to 0.076. Heritability for survival days (SL) was found to be closer with that of binary records (S) at each time of observation. While slightly higher heritability estimates were found for S than for SL at early stages of the grow-out testing period in the BL2019 line, they were very closer in the middle to later test periods. The trajectories of S and

TABLE 3 Estimates of variance components and heritabilities ($h^2 \pm se$) for the genetic analyse of survival data and body weight.

| Trait | *BL2019* | | | | *BL2020* | | | |
|---|---|---|---|---|---|---|---|---|
| | $\sigma_\alpha^2$ | $\sigma_p^2$ | $\sigma_e^2$ | $h^2 \pm se$ | $\sigma_\alpha^2$ | $\sigma_p^2$ | $\sigma_e^2$ | $h^2 \pm se$ |
| S1 | 1.09 $E^{-5}$ | 2.31 $E^{-3}$ | 2.30 $E^{-3}$ | 0.005 ± 0.002 | 3.02 $E^{-4}$ | 1.53 $E^{-2}$ | 1.50 $E^{-2}$ | 0.020 ± 0.004 |
| S2 | 6.64 $E^{-5}$ | 3.83 $E^{-3}$ | 3.76 $E^{-3}$ | 0.017 ± 0.003 | 5.88 $E^{-4}$ | 1.86 $E^{-2}$ | 1.80 $E^{-2}$ | 0.017 ± 0.003 |
| S3 | 2.11 $E^{-4}$ | 6.21 $E^{-3}$ | 6.00 $E^{-3}$ | 0.034 ± 0.005 | 1.16 $E^{-3}$ | 2.70 $E^{-2}$ | 2.58 $E^{-2}$ | 0.043 ± 0.006 |
| S4 | 4.59 $E^{-4}$ | 9.99 $E^{-3}$ | 9.53 $E^{-3}$ | 0.046 ± 0.005 | 2.20 $E^{-3}$ | 3.96 $E^{-2}$ | 3.74 $E^{-2}$ | 0.056 ± 0.007 |
| S5 | 9.57 $E^{-4}$ | 1.58 $E^{-2}$ | 1.49 $E^{-2}$ | 0.060 ± 0.007 | 3.11 $E^{-3}$ | 5.08 $E^{-2}$ | 4.77 $E^{-2}$ | 0.061 ± 0.007 |
| S6 | 1.27 $E^{-3}$ | 2.03 $E^{-2}$ | 1.90 $E^{-2}$ | 0.063 ± 0.008 | 4.38 $E^{-3}$ | 6.32 $E^{-2}$ | 5.89 $E^{-2}$ | 0.069 ± 0.008 |
| S7 | 1.45 $E^{-3}$ | 2.62 $E^{-2}$ | 2.48 $E^{-2}$ | 0.055 ± 0.007 | 6.31 $E^{-3}$ | 9.06 $E^{-2}$ | 8.43 $E^{-2}$ | 0.070 ± 0.008 |
| S8 | 2.08 $E^{-3}$ | 3.49 $E^{-2}$ | 3.29 $E^{-2}$ | 0.060 ± 0.007 | 8.21 $E^{-3}$ | 1.09 $E^{-1}$ | 1.00 $E^{-1}$ | 0.076 ± 0.009 |
| S9 | 3.16 $E^{-3}$ | 5.38 $E^{-2}$ | 5.07 $E^{-2}$ | 0.059 ± 0.007 | 8.88 $E^{-3}$ | 1.17 $E^{-1}$ | 1.08 $E^{-1}$ | 0.076 ± 0.009 |
| S10 | 4.85 $E^{-3}$ | 7.11 $E^{-2}$ | 6.62 $E^{-2}$ | 0.068 ± 0.008 | 9.15 $E^{-3}$ | 1.21 $E^{-1}$ | 1.12 $E^{-1}$ | 0.075 ± 0.009 |
| S11 | 5.32 $E^{-3}$ | 8.08 $E^{-2}$ | 7.55 $E^{-2}$ | 0.066 ± 0.008 | *NA* | *NA* | *NA* | *NA* |
| S12 | 5.86 $E^{-3}$ | 8.70 $E^{-2}$ | 8.12 $E^{-2}$ | 0.067 ± 0.008 | *NA* | *NA* | *NA* | *NA* |
| SL1 | 1.14 $E^{-4}$ | 1.10 $E^{-1}$ | 1.10 $E^{-1}$ | 0.001 ± 0.002 | 1.09 $E^{-2}$ | 6.34 $E^{-1}$ | 6.23 $E^{-1}$ | 0.017 ± 0.003 |
| SL2 | 3.79 $E^{-3}$ | 6.56 $E^{-1}$ | 6.52 $E^{-1}$ | 0.006 ± 0.002 | 9.25 $E^{-2}$ | 4.17 $E^{0}$ | 4.08 $E^{0}$ | 0.022 ± 0.004 |
| SL3 | 2.58 $E^{-2}$ | 1.96 $E^{0}$ | 1.93 $E^{0}$ | 0.013 ± 0.003 | 3.29 $E^{-1}$ | 1.16 $E^{1}$ | 1.13 $E^{1}$ | 0.029 ± 0.005 |
| SL4 | 1.11 $E^{-1}$ | 4.46 $E^{0}$ | 4.35 $E^{0}$ | 0.025 ± 0.004 | 9.16 $E^{-1}$ | 2.46 $E^{1}$ | 2.37 $E^{1}$ | 0.037 ± 0.005 |
| SL5 | 3.41 $E^{-1}$ | 8.93 $E^{0}$ | 8.58 $E^{0}$ | 0.038 ± 0.005 | 2.06 $E^{0}$ | 4.55 $E^{1}$ | 4.35 $E^{1}$ | 0.045 ± 0.006 |
| SL6 | 8.28 $E^{-1}$ | 1.65 $E^{1}$ | 1.56 $E^{1}$ | 0.050 ± 0.007 | 4.01 $E^{0}$ | 7.63 $E^{1}$ | 7.23 $E^{1}$ | 0.052 ± 0.007 |
| SL7 | 1.61 $E^{0}$ | 2.80 $E^{1}$ | 2.63 $E^{1}$ | 0.058 ± 0.007 | 7.05 $E^{0}$ | 1.19 $E^{2}$ | 1.12 $E^{2}$ | 0.059 ± 0.007 |
| SL8 | 2.80 $E^{0}$ | 4.45 $E^{1}$ | 4.17 $E^{1}$ | 0.063 ± 0.008 | 1.16 $E^{1}$ | 1.79 $E^{2}$ | 1.67 $E^{2}$ | 0.065 ± 0.008 |
| SL9 | 4.50 $E^{0}$ | 6.78 $E^{1}$ | 6.33 $E^{1}$ | 0.066 ± 0.008 | 1.79 $E^{1}$ | 2.59 $E^{2}$ | 2.41 $E^{2}$ | 0.069 ± 0.008 |
| SL10 | 6.98 $E^{0}$ | 1.01 $E^{2}$ | 9.37 $E^{1}$ | 0.069 ± 0.008 | 2.60 $E^{1}$ | 3.61 $E^{2}$ | 3.35 $E^{2}$ | 0.072 ± 0.008 |
| SL11 | 1.05 $E^{1}$ | 1.47 $E^{2}$ | 1.36 $E^{2}$ | 0.071 ± 0.008 | *NA* | *NA* | *NA* | *NA* |
| SL12 | 1.50 $E^{1}$ | 2.08 $E^{2}$ | 1.93 $E^{2}$ | 0.072 ± 0.008 | *NA* | *NA* | *NA* | *NA* |
| SUR | 1.45 $E^{-2}$ | 1.84 $E^{-1}$ | 1.70 $E^{-1}$ | 0.079 ± 0.008 | 3.86 $E^{-2}$ | 4.89 $E^{-1}$ | 4.50 $E^{-1}$ | 0.079 ± 0.003 |
| BW | 1.59 $E^{0}$ | 3.34 $E^{1}$ | 1.75 $E^{1}$ | 0.486 ± 0.036 | 6.69 $E^{0}$ | 1.79 $E^{1}$ | 1.12 $E^{1}$ | 0.373 ± 0.031 |

*Note:* $\sigma_\alpha^2$, $\sigma_p^2$, *and* $\sigma_e^2$ represent additive genetic variance, phenotype variance, and residual variance; S1-S12, SL1-SL12, SUR, and BW: see legend in Table 2; *NA*, not applicable.

SL during grow-out test period are illustrated in Figure 2. The slightly higher heritability estimates for survival in BL2020 than in BL2019 were expected since we recorded a higher daily mortality in BL 2020. It was interesting that, for both 2-year data groups, heritability of S showed a steady increase over time. While SL estimates increased during the first 60 days, levels of heritability for SL fluctuated at ~0.06 and maintained a plateau after this initial period.

Heritability estimates for final population census survival (SUR) were very similar for S and SL at the final testing stage, while for body weight (BW) in BL2019 and BL 2020, they were 0.486 ± 0.036 and 0.373 ± 0.031, respectively.

## Phenotype and genetic correlations

Summaries of phenotype ($r_p$) and genetic ($r_g$) correlations with *SE* for S, SL, SUR, and BW are provided in Supplementary Tables S1–S4. In general $r_g$ for S, estimates between different times were moderate to high and ranged from 0.475 (S1 *vs* S12) to 0.999 (S11 *vs* S12) in the BL2019 breeding line, and 0.277 (S1 *vs* S10) to 0.999 (S9 *vs* S10) for BL 2020, respectively. Estimates of $r_g$ for SL between different times were much closer compared with that of S in both 2-year breeding lines. As expected, estimates of $r_g$ between SUR and S, or SUR and SL were much lower, and in most cases were only moderate correlations likely due to the issues referred to unobserved mortality events when estimating SUR vs. S/SL. In contrast, estimates of $r_g$ for BW and all survival related traits (S/SL/SUR) were low but all positive. We consider that this result implies that selection directed on survival related traits would not experience potentially negative tradeoff effects on growth traits. Patterns of phenotype correlations ($r_p$) among S, SL, SUR, and BW were similar to those for $r_g$ but were always slightly lower than for $r_g$ (Figure 3).

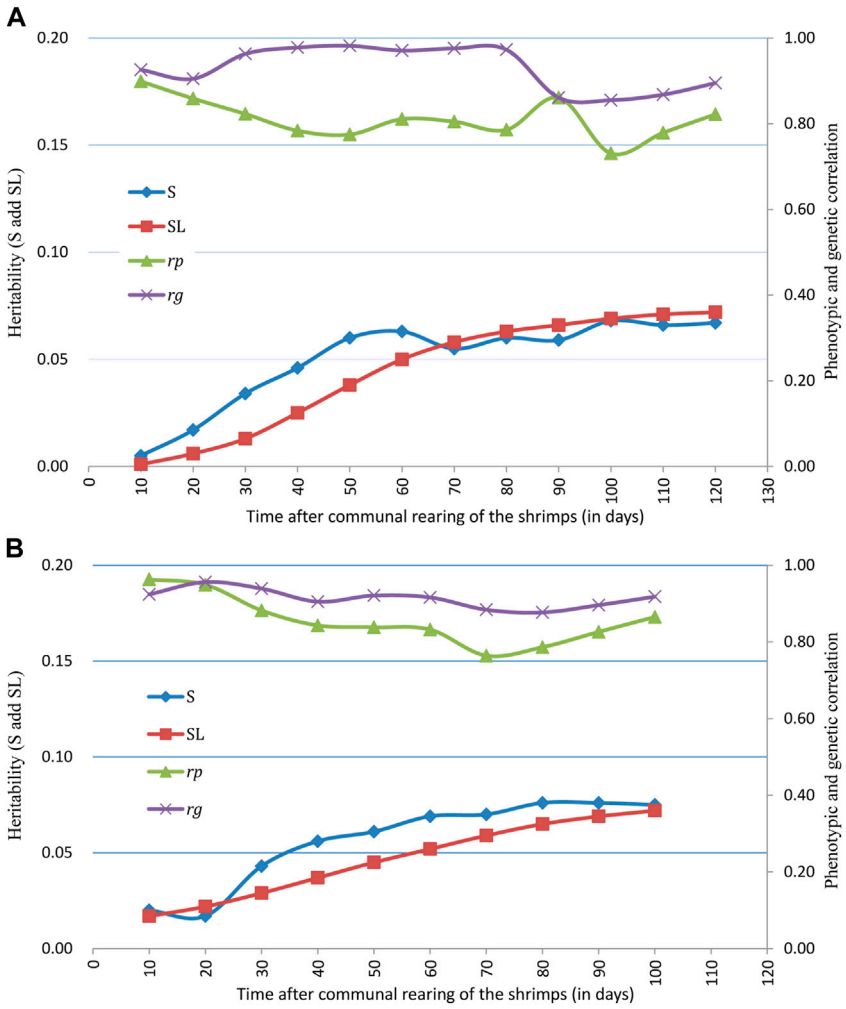



FIGURE 2
Trajectories of heritability for survival traits (S and SL), phenotypic correlation ($r_p$) and genetic correlation ($r_g$) between S and SL across grow-out test period of **(A)** BL2019 line, **(B)** BL2020 line.

Trajectory of correlations between S and SL estimates across the grow-out test period are illustrated in Figure 2. Genetic correlations between binary survival traits (S) and survival days (SL) were all high at the different observation times and ranged from 0.855 to 0.982 for BL2019 and 0.877 to 0.956 for BL 2020. The results reported above suggest that families with high survival rates are also likely to show a general trend for a longer mean life span over the test period. Trajectories of genetic correlations for S traits across time were consistent with that of estimation for SL, with estimates for $r_g$ gradually decreasing across the experimental time period (Figure 3). Patterns of change on $r_g$ estimates for S/SL across time were similar between breeding lines in both years.

## Full-sib family effects

The likelihood ratio test suggested that there were very limited full-sib family effects ($c^2$) in the genetic analysis. Among 26 comparison for BL2019, 24 tests were not significant for the likelihood ratio tests (Supplementary Table S5). In contrast, S1 and SUR were significant for $c^2$ ($\chi^2_{1df}$). Data interpretation for genetic parameters for S1 and SUR however, showed limited potential for improving an animal model fit for full-sib family effects (Model 3) because the new results for S1 and SUR genetic parameters showed large associated SE estimates (S1, $h^2$ = 0.005 ± 0.009; SUR, $h^2$ = 0.053 ± 0.070). Similarly, there was limited potential for improving animal model fit for full-sib family effects in the genetic analysis of survival data from BL2020 based on the result of likelihood ratio tests.

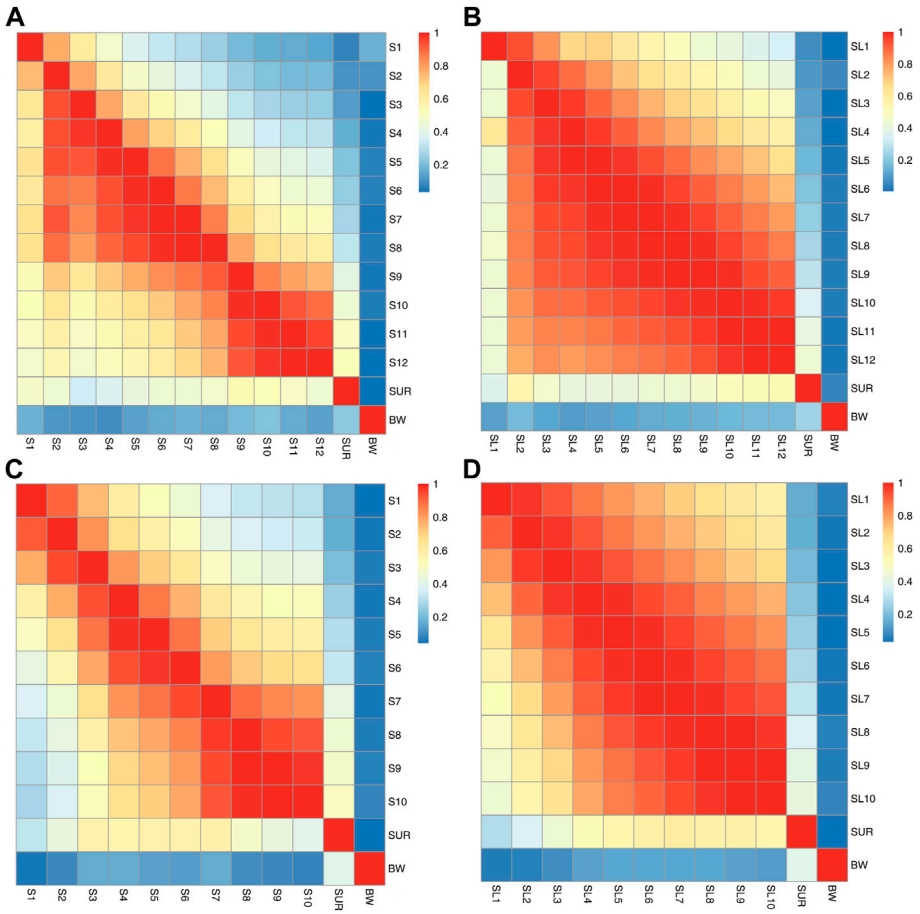



FIGURE 3
The trajectory of phenotype (above diagonal) and genetic correlations (below diagonal) between **(A)** S1-S12, SUR and BW for the breeding line of BL 2019, **(B)** SL1-SL12, SUR and BW of BL 2019, **(C)** S1-S10, SUR and BW for the breeding line of BL 2020, **(D)** SL1-SL10, SUR and BW for the breeding line of BL 2020.

# Discussion

The current study provides the first genetic analysis of the trajectory of survival traits across grow-out stage for improved lines in a farmed aquatic species. Results of the genetic evaluation indicate that heritability estimates for the two defined types survival applied here were very similar. In general, estimated levels of heritability for survival traits were low but increased gradually across the grow-out test periods for both year classes. In addition, genetic correlations between the binary trait of survivability (S) and the continuous trait of survival time (SL) were relatively high. These findings will be useful for genetic analyses of survival data in field environments and potentially contribute to the development of relatively robust farmed aquatic strains used in aquaculture. An issue we identified in the analysis however, was missing records of some individual mortality events across our test periods. This reflects in general the difficulty of collecting survival data for aquaculture species in the field when working in commercial test environments.

## Mortality

Records of daily mortality in breeding lines are not only important for genetic evaluation of survival traits, but also for assessment of general husbandry management of population health status in breeding lines. Changes in daily mortality here were similar to those reported in our previous study with the mortality rate varying between 0.1% and 0.15% (Ren et al., 2020a). Designing and managing reliable, high quality water culture conditions is a major constraint in selective breeding programs for penaeid shrimps. Most daily mortality rates for farmed penaeids domestication programs range between 0.1% and 0.5%, depending on the different types of culture systems employed (Yano, 2000; Coman et al., 2005; Duy et al., 2012).

The highest daily mortality peaks in both test years occurred in late grow-out stages (Figure 1), periods that coincided with the coldest weather temperatures in the winter seasons at the hatchery location in China. In general, husbandry

management of penaeid breeding lines are considered appropriate if daily mortality rate fluctuates around 0.5%, and higher mortality rates do not last for longer than 7–10 days. In our study, when mean daily temperatures increased from this relatively low range, mortality decreased to normal levels again. Furthermore, in most cases individuals that died during low temperature periods were of smaller or weaker individuals.

In the current study, we observed a gap in bias between survival estimators across the test period and the final population census data. Collecting survival data in natural 'field' conditions can be difficult in aquaculture breeding programs compared with that in terrestrial farm animals because most mortality events are under water and consequently, may not be detected. While survival phenotype derived from controlled experimental challenge tests can be more accurate (Ødegård et al., 2011; Yáñez et al., 2014), the use of survival data in natural field conditions can often provide better information for developing robust culture lines in aquaculture species (Houston et al., 2008; Lillehammer et al., 2013; Bangera et al., 2014; Dégremont et al., 2015). In particular, survival data collected under field conditions will better reflect natural processes of changes in mortality in a population and collecting these data can avoid potential G-by-E impacts when using artificial challenge test environments.

Survival estimators and final population census results in the current study were comparable with earlier reports on genetic breeding programs conducted on penaeid shrimps. In Mexico, survival rate of *P. vannamei* selection lines were reported to range from 71% to 82.2% during the grow-out test stages (Campos-Montes et al., 2013; Caballero-Zamora et al., 2015). Breeding nucleus data for farmed *P. vannamei* in Colombia reported survival rates ranging from 56.9% to 77.2% (Gitterle et al., 2005), while they were 70% and 73% for two grow-out stages in a selection program for *P. monodon* in Australia (Coman et al., 2010). In Vietnam, survival rates for *P. monodon* were 34%–49% in a family selection program (Van Sang et al., 2020).

## Animal models for survival analysis

While survival traits are often recorded as non-normally distributed binary traits, reports of genetic evaluation of survival data in dairy cattle (Sasaki et al., 2015; Van Pelt et al., 2015, 2016; Heise et al., 2016, 2018) suggest that linear models are likely to be more appropriate for genetic analysis of survival data in aquaculture. Proportion hazards models (PHM) have been often considered to fit time-to-event survival data better (Ducrocq, 1994; Neerhof et al., 2000), however, computing time to analyse these data successfully is much high than with linear models. Furthermore, using a PHM approach, it is difficult to assess genetic correlations with other continuous traits. Similarly, threshold models require significantly more computing resources for genetic analysis of survival data compared with linear models. Moreover, correlations of EBVs estimated with threshold models and linear models have been almost identical, indicating that there are few advantages of applying threshold models in routine genetic analyses of survival data (Boettcher et al., 1999; Vazquez et al., 2009). In practical terms, consideration of model choice in breeding programs can be a balance between multiple factors, including: simplified data records, comparative model performance, available computing resources, and integration of data from other important economic traits at the same time. We therefore consider linear multivariate animal models (LMA) to provide informative and reliable tools for routine genetic analysis of survival data in farmed aquatic species.

Data structure in the current study applied a larger number of full paternity and half-sib families that were suitable for achieving highly accurate heritability estimates. In addition, full and half-sib families were produced over a relatively short time period (1 week) that effectively removes time of age effects and common environmental effects in the genetic analysis. This can enhance the accuracy of the animal model results generated. We therefore chose a linear multivariate animal model (LMA) approach for our genetic analyses rather than other more sophisticated linear animal models such as a random regression model or repeatability model to fit age effects of time for the survival data set here.

## Heritability estimates

Heritability estimates for S1 to S12 and SL1 to SL 12 were consistent with results reported from other pacific white shrimp breeding programs. Gitterle et al. (2005) evaluated binary survival traits based on 430 full-sib families in different test environments and reported heritability estimates that ranged from 0.04 to 0.10. In another genetic analysis of 2008–2010 survival data of pacific white shrimp in Mexico, heritability estimates for early stage survival were 0.03 and 0.04 for later grow-out test stage (Campos-Montes et al., 2013). In a G-by-E effect study, heritability of survival was reported to be 0.06 under cold temperature conditions, while it was much higher (0.11) at normal ambient temperatures (Li et al., 2015). In addition, genetic parameters for survival in the presence or absence of a white spot disease outbreak were reported to be 0.00 and 0.06 (Caballero-Zamora et al., 2015), respectively. Overall, these results for heritability estimates of survival traits agree well with quantitative genetic theoretical predictions that survival traits are a group of fitness-associated traits that tend to show the lowest levels of heritability (Hill, 2010).

Here we compared two types of survival definitions; binary traits and continuous traits for our genetic evaluation, with both showing similar heritability estimates (Table 3). While this result did not fit our earlier assumption that continuous traits would

improve the results for heritability estimates *via* adding more survival time information to mortality events, similar findings have been reported for genetic analysis of survival data in experimental challenge tests on aquaculture species (Joshi et al., 2021a; Vu et al., 2022). Binary trait recording however, is much more simple (0, 1) to score than continuous survival data in practice and thus, should be considered more feasible for routine genetic evaluation in aquaculture breeding programs.

Knowledge about heritability estimate trajectories across time is important to improve efficiency in selection programs and to provide insights on predicted selection response over the time period in the test. We believe that our study is the first report on the heritability of survival trajectory across grow-out stage in farmed aquatic species. Results here, suggest that for both lines and applying both types of survival trait definition, similar patterns for trajectory with time were evident. In general, initial heritability estimates were almost close to 0, following which they gradually increased across the grow-out test period before slowing down or reaching a plateau during the latter stages (Figure 2). Of interest, this pattern in trajectory is similar compared with some earlier reports for changes in growth heritability estimates across different age times, whereas heritability estimates for growth traits reached a plateau phase in the middle to later grow-out testing stage (Turra et al., 2012; He et al., 2017). In dairy science, similar patterns for survival trait heritability were also reported for a 72 month period of milk production life (Van Pelt et al., 2015). Therefore, this pattern of heritability estimate trajectories across time can be very useful for making critical decisions in a breeding plan.

Common full-sib effects ($c^2$) were not significant in the current study indicating that linear multivariate animal model (LMA) without full-sib effects was effective for routine genetic evaluation here. This finding agrees with a systematic review paper on $c^2$ estimation in aquaculture breeding suggests that full-sib effects contribute only a small proportion of total phenotypic variance at earlier growth stages for growth related traits, but for growth traits in an individual's later stages or with other phenotypic traits, full-sib effects ($c^2$) are not significant in most cases and are essentially zero (Nguyen, 2021). As an example, routine genetic evaluation models applied in commercial tilapia selection breeding programs based on current published genetic animal models do not include full-sib effects ($c^2$) (Joshi et al., 2021b).

## Genetic and phenotype correlations

Both genetic and phenotypic correlations between binary and continuous survival traits time showed high correlations (>0.8) at each time window (Figure 2). In an aquaculture context, similar patterns have been reported for correlations between different types of growth related traits, including for $r_g/r_p$ among body weight, body length, and other morphological growth traits (He et al., 2017; Schlicht et al., 2018). Patterns for $r_g/r_p$ also support findings from heritability estimates that both survival definitions can be applied equally well for genetic selection on overall survival trait data in commercial breeding lines.

The current study is the first report on patterns of genetic correlations for survival traits across grow-out testing stage in a farmed aquatic species (Figure 3). In general, patterns of genetic correlation for survival traits across time ($r_g$ among S1 to S12 and SL1 to SL12) in the current study were moderate to high. There was a trend here for a gradual decrease in $r_g$ between survival traits with the time. These patterns for $r_g$ were in accordance with trajectories of genetic correlations for growth traits across ages in other aquaculture studies (Turra et al., 2012; He et al., 2017; Schlicht et al., 2018), as well as for survival traits in livestock (Van Pelt et al., 2015). In addition, results for $r_p$ trajectories of correlation across time between S1 to S12 and/or SL1 to SL12 were similar with that of $r_g$, but values for $r_p$ are always slightly lower than for $r_g$ (Figure 3).

In contrast, genetic correlations between SUR and S1-12/SL1-12 were much lower but nevertheless, in most instances were still moderate (Figure 3). From observation of the patterns of survival traits across times, the real genetic correlations between SUR and other survival traits are likely to be much higher than that observed here. This most likely resulted from the missing data of survival records during the grow-out period and the final population census. Highly accurate recording of survival data in field commercial aquaculture environments however, is often difficult compared with data collection in challenge test experimental conditions, particularly, for pacific white shrimp that are of relatively small size and that are cannibalistic. These characteristics mean that some results of mortality events during grow-out test period are unlikely ever to be detected under field conditions.

In the current study, correlations between body weight and survival traits were generally low (Figure 3), but still positive, a result that is consistent with previous work on pacific white shrimp (Campos-Montes et al., 2013; Caballero-Zamora et al., 2015; Li et al., 2015; Ren et al., 2020a). This suggests overall, that survival and growth are two groups of separate traits and can be selected for simultaneously in a genetic breeding program *via* a multi-traits selection approach.

## Implications

While survival can be considered a relatively 'old' trait in animal breeding, often neglected because of the simple records (0 or 1) scored and generally low heritability estimates. Developing routine genetic evaluation systems for survival traits however, will be important for developing robust strains using valuable field survival data. Firstly, survival data collection from commercial lines can be used directly for improving overall survival rate in target breeding lines (Gjedrem and Rye, 2018).

Moreover, survival data recorded from commercial lines when special events arise e.g., disease outbreaks can also be applied effectively to developing disease resistant strains, and these type of data are also valuable for avoiding potential G-by-E effects compared with survival data from controlled (artificial) challenge experiments (Bangera et al., 2014; Dégremont et al., 2015; Barría et al., 2020; Fraslin et al., 2022). Additionally, despite of low heritability of survival traits at the early grow-out testing stage, routine genetic evaluation of survival data in breeding lines provides a vital assessment of the relative health status of a stock under the general husbandry management practices employed in the breeding lines. Here we provide the first report of the trajectory of heritability estimates for survival traits across the grow-out period for pacific white shrimp breeding lines in China. The results will be useful for applying commercial field survival data to develop improved robust culture lines of this important crustacean species in the future.

## Conclusions

In conclusion, we developed routine genetic evaluations for survival data in selective breeding pacific white shrimp line in China. Results of heritability and genetic correlation estimates indicate that both survival definitions for binary traits and continuous traits of survival time can be used effectively for genetic analysis of survival data. Binary survival records following with linear multivariate animal models (LMA) provide a feasible routine genetic evaluation approach in practical commercial breeding programs because of the simple data recording, reduced computing time required, and the ability to combine performance assessment of multiple continuous phenotypic traits in the genetic analysis. While heritability estimates for survival traits here were generally low, they showed a gradual increasing trend across the grow-out period. Genetic correlations for survival traits during grow-out tests were moderate to high, and the closer the times were between estimates of two survival traits, the higher were the genetic correlations. Genetic correlations between survival traits and body weight were low but all estimates were positive. In summary, we first reported the trajectory of heritability estimates for survival traits across grow-out stage in aquaculture genetics, which would be useful for how to use survival data in field situations to develop more robust culture strains of *P. vannamei* and potentially for other farmed aquatic species in the future.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## Author contributions

SR and DH conceptualized the research, developed animal models for data analyses. SR and PM wrote the manuscript. All authors have critically revised the manuscript.

## Acknowledgments

The authors would like to thank Haixiang Lan for collecting the survival data used in this study and thank Zhante Shi for technical support during the grow-out testing of the shrimp breeding lines.

## Conflict of interest

BT is employed by the Company Beijing Shuishiji Biotechnology Co., Ltd.

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1018568/full#supplementary-material

## References

Bangera, R., Ødegård, J., Mikkelsen, H., Nielsen, H. M., Seppola, M., Puvanendran, V., et al. (2014). Genetic analysis of francisellosis field outbreak in Atlantic cod (*Gadus morhua* L.) using an ordinal threshold model. *Aquaculture* 420, S50–S56. doi:10.1016/j.aquaculture.2013.08.029

Barría, A., Christensen, K. A., Yoshida, G., Jedlicki, A., Leong, J. S., Rondeau, E. B., et al. (2019). Whole genome linkage disequilibrium and effective population size in a coho salmon (*Oncorhynchus kisutch*) breeding population using a high-density SNP array. *Front. Genet.* 10, 498. doi:10.3389/fgene.2019.00498

Barría, A., Trinh, T. Q., Mahmuddin, M., Benzie, J. A., Chadag, V. M., and Houston, R. D. (2020). Genetic parameters for resistance to Tilapia lake virus (TiLV) in nile tilapia (*Oreochromis niloticus*). *Aquaculture* 522, 735126. doi:10.1016/j.aquaculture.2020.735126

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M., Beacham, T. D., et al. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends Ecol. Evol.* 32 (9), 665–680. doi:10.1016/j.tree.2017.06.010

Boettcher, P. J., Jairath, L. K., and Dekkers, J. C. M. (1999). Comparison of methods for genetic evaluation of sires for survival of their daughters in the first three lactations. *J. Dairy Sci.* 82 (5), 1034–1044. doi:10.3168/jds.S0022-0302(99)75324-5

Caballero-Zamora, A., Montaldo, H. H., Campos-Montes, G. R., Cienfuegos-Rivas, E. G., Martínez-Ortega, A., and Castillo-Juarez, H. (2015). Genetic parameters for body weight and survival in the Pacific White Shrimp *Penaeus (Litopenaeus) vannamei* affected by a White Spot Syndrome Virus (WSSV) natural outbreak. *Aquaculture* 447, 102–107. doi:10.1016/j.aquaculture.2014.08.028

Campos-Montes, G. R., Montaldo, H. H., Martínez-Ortega, A., Jiménez, A. M., and Castillo-Juarez, H. (2013). Genetic parameters for growth and survival traits in Pacific white shrimp *Penaeus (Litopenaeus) vannamei* from a nucleus population undergoing a two-stage selection program. *Aquac. Int.* 21 (2), 299–310. doi:10.1007/s10499-012-9553-1

Coman, G. I., Crocos, P. I., Arnold, S. J., Keys, S. I., Preston, N. P., and Murphy, B. (2005). Growth, survival and reproductive performance of domesticated Australian stocks of the giant tiger prawn, *Penaeus monodon*, reared in tanks and raceways. *J. World Aquac. Soc.* 36 (4), 464–479. doi:10.1111/j.1749-7345.2005.tb00394.x

Coman, G. J., Arnold, S. J., Wood, A. T., and Kube, P. D. (2010). Age: Age genetic correlations for weight of *Penaeus monodon* reared in broodstock tank systems. *Aquaculture* 307 (1-2), 1–5. doi:10.1016/j.aquaculture.2010.06.027

Correa, K., Lhorente, J. P., López, M. E., Bassini, L., Naswa, S., Deeb, N., et al. (2015). Genome-wide association analysis reveals loci associated with resistance against Piscirickettsia salmonis in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC Genomics* 16 (1), 854–859. doi:10.1186/s12864-015-2038-7

Dégremont, L., Garcia, C., and Allen, S. K., Jr (2015). Genetic improvement for disease resistance in oysters: A review. *J. Invertebr. Pathol.* 131, 226–241. doi:10.1016/j.jip.2015.05.010

Ducrocq, V., and Casella, G. (1996). A Bayesian analysis of mixed survival models. *Genet. Sel. Evol.* 28 (6), 505–529. doi:10.1186/1297-9686-28-6-505

Ducrocq, V., Sölkner, J., and Mészáros, G. (2010). *Survival Kit V6—a software package for survival analysis*. Leipzig, Germany: 9th World Congr. Genet. Appl. Livest. Prod.

Ducrocq, V., and Sölkner, J. (1994). "The survival kit": A fortran package for the analysis of survival data," in *Proc.5th world congr. Genet. Appl. Livest. Prod.* Guelph, Ontario, Canada, 51–52.

Ducrocq, V. (1994). Statistical analysis of length of productive life for dairy cows of the Normande breed. *J. Dairy Sci.* 77 (3), 855–866. doi:10.3168/jds.S0022-0302(94)77020-X

Duy, H. N., Coman, G. J., Wille, M., Wouters, R., Quoc, H. N., Vu, T., et al. (2012). Effect of water exchange, salinity regime, stocking density and diets on growth and survival of domesticated black tiger shrimp *Penaeus monodon* (Fabricius, 1798) reared in sand-based recirculating systems. *Aquaculture* 338, 253–259. doi:10.1016/j.aquaculture.2012.01.021

FAO (2020). *Aquaculture production (quanlities and values). FishStatJ -software for fishery statistical time series*. Rome: FAO Fisheries and Aquaculture Department.

Forabosco, F., Jakobsen, J. H., and Fikse, W. F. (2009). International genetic evaluation for direct longevity in dairy bulls. *J. Dairy Sci.* 92 (5), 2338–2347. doi:10.3168/jds.2008-1214

Friggens, N. C., Blanc, F., Berry, D. P., and Puillet, L. (2017). Review: Deciphering animal robustness. A synthesis to facilitate its use in livestock breeding and management. *Animal* 11 (12), 2237–2251. doi:10.1017/S175173111700088X

Gitterle, T., Rye, M., Salte, R., Cock, J., Johansen, H., Lozano, C., et al. (2005). Genetic (co) variation in harvest body weight and survival in *Penaeus (Litopenaeus) vannamei* under standard commercial conditions. *Aquaculture* 243 (1-4), 83–92. doi:10.1016/j.aquaculture.2004.10.015

Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* 350, 117–129. doi:10.1016/j.aquaculture.2012.04.008

Gjedrem, T., and Rye, M. (2018). Selection response in fish and shellfish: A review. *Rev. Aquac.* 10 (1), 168–179. doi:10.1111/raq.12154

Gutierrez, A. P., Bean, T. P., Hooper, C., Stenton, C. A., Sanders, M. B., Paley, R. K., et al. (2018). A genome-wide association study for host resistance to ostreid herpesvirus in Pacific oysters (*Crassostrea gigas*). *G3-Genes Genom. Genet.* 8 (4), 1273–1280. doi:10.1534/g3.118.200113

He, J., Zhao, Y., Zhao, J., Gao, J., Han, D., Xu, P., et al. (2017). Multivariate random regression analysis for body weight and main morphological traits in genetically improved farmed tilapia (*Oreochromis niloticus*). *Genet. Sel. Evol.* 49 (1), 80–13. doi:10.1186/s12711-017-0357-7

Heise, J., Liu, Z., Stock, K. F., Rensing, S., Reinhardt, F., and Simianer, H. (2016). The genetic structure of longevity in dairy cows. *J. Dairy Sci.* 99 (2), 1253–1265. doi:10.3168/jds.2015-10163

Heise, J., Stock, K. F., Reinhardt, F., Ha, N. T., and Simianer, H. (2018). Phenotypic and genetic relationships between age at first calving, its component traits, and survival of heifers up to second calving. *J. Dairy Sci.* 101 (1), 425–432. doi:10.3168/jds.2017-12957

Hill, W. G. (2010). Understanding and using quantitative genetic variation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365 (1537), 73–85. doi:10.1098/rstb.2009.0203

Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., et al. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nat. Rev. Genet.* 21 (7), 389–409. doi:10.1038/s41576-020-0227-y

Houston, R. D. (2017). Future directions in breeding for disease resistance in aquaculture species. *R. Bras. Zootec.* 46, 545–551. doi:10.1590/s1806-92902017000600010

Houston, R. D., Haley, C. S., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., et al. (2008). Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* 178 (2), 1109–1115. doi:10.1534/genetics.107.082974

Hung, D., Nguyen, N. H., Hurwood, D. A., and Mather, P. B. (2013). Quantitative genetic parameters for body traits at different ages in a cultured stock of giant freshwater prawn (*Macrobrachium rosenbergii*) selected for fast growth. *Mar. Freshw. Res.* 65 (3), 198–205. doi:10.1071/mf13111

Joshi, R., Skaarud, A., Alvarez, A. T., Moen, T., and Ødegård, J. (2021b). Bayesian genomic models boost prediction accuracy for survival to *Streptococcus agalactiae* infection in Nile tilapia (*Oreochromus nilioticus*). *Genet. Sel. Evol.* 53 (1), 37–10. doi:10.1186/s12711-021-00629-y

Joshi, R., Skaaurd, A., and Tola Alvarez, A. (2021a). Experimental validation of genetic selection for resistance against Streptococcus agalactiae via different routes of infection in the commercial Nile tilapia breeding programme. *J. Anim. Breed. Genet.* 138 (3), 338–348. doi:10.1111/jbg.12516

Kaplan, E. L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* 53 (282), 457–481. doi:10.1080/01621459.1958.10501452

Kassambara, A., Kosinski, M., Biecek, P., and Fabian, S. (2017). Package 'survminer'. Drawing Survival Curves using 'ggplot2'. R package version 0.3.1.

Kumar, G., and Engle, C. R. (2016). Technological advances that led to growth of shrimp, salmon, and tilapia farming. *Rev. Fish. Sci. Aquac.* 24 (2), 136–152. doi:10.1080/23308249.2015.1112357

Li, W., Luan, S., Luo, K., Sui, J., Xu, X., Tan, J., et al. (2015). Genetic parameters and genotype by environment interaction for cold tolerance, body weight and survival of the Pacific white shrimp *Penaeus vannamei* at different temperatures. *Aquaculture* 441, 8–15. doi:10.1016/j.aquaculture.2015.02.013

Lillehammer, M., Ødegård, J., Madsen, P., Gjerde, B., Refstie, T., and Rye, M. (2013). Survival, growth and sexual maturation in atlantic salmon exposed to infectious pancreatic necrosis: A multi-variate mixture model approach. *Genet. Sel. Evol.* 45 (1), 8–12. doi:10.1186/1297-9686-45-8

Meyer, K. (2007). WOMBAT—a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B* 8 (11), 815–821. doi:10.1631/jzus.2007.B0815

Naylor, R. L., Hardy, R. W., Buschmann, A. H., Bush, S. R., Cao, L., Klinger, D. H., et al. (2021). A 20-year retrospective review of global aquaculture. *Nature* 591 (7851), 551–563. doi:10.1038/s41586-021-03308-6

Neerhof, H. J., Madsen, P., Ducrocq, V. P., Vollema, A. R., Jensen, J., and Korsgaard, I. R. (2000). Relationships between mastitis and functional longevity in Danish Black and White dairy cattle estimated using survival analysis. *J. Dairy Sci.* 83 (5), 1064–1071. doi:10.3168/jds.S0022-0302(00)74970-8

Nguyen, N. H. (2021). A systematic review and meta-analysis of genetic parameters for complex quantitative traits in aquatic animal species. *bioRxiv*. doi:10.1101/2021.05.20.445048

Nguyen, N. H. (2016). Genetic improvement for important farmed aquaculture species with a reference to carp, tilapia and prawns in asia: Achievements, lessons and challenges. *Fish. Fish.* 17 (2), 483–506. doi:10.1111/faf.12122

Ødegård, J., Baranski, M., Gjerde, B., and Gjedrem, T. (2011). Methodology for genetic evaluation of disease resistance in aquaculture species: Challenges and future prospects. *Aquac. Res.* 42, 103–114. doi:10.1111/j.1365-2109.2010.02669.x

Ødegård, J., Olesen, I., Gjerde, B., and Klemetsdal, G. (2006). Evaluation of statistical models for genetic analysis of challenge test data on furunculosis resistance in Atlantic salmon (*Salmo salar*): Prediction of field survival. *Aquaculture* 259 (1-4), 116–123. doi:10.1016/j.aquaculture.2006.05.034

Palaiokostas, C., Cariou, S., Bestin, A., Bruant, J. S., Haffray, P., Morin, T., et al. (2018). Genome-wide association and genomic prediction of resistance to viral nervous necrosis in European sea bass (*Dicentrarchus labrax*) using RAD sequencing. *Genet. Sel. Evol.* 50 (1), 30–11. doi:10.1186/s12711-018-0401-2

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Core Team.

Reid, G. K., Gurney-Smith, H. J., Marcogliese, D. J., Knowler, D., Benfey, T., Garber, A. F., et al. (2019). Climate change and aquaculture: Considering biological response and resources. *Aquac. Environ. Interact.* 11, 569–602. doi:10.3354/aei00332

Ren, S., Mather, P. B., Prentis, P., Li, Y., Tang, B., and Hurwood, D. A. (2020b). Quantitative genetic assessment of female reproductive traits in a domesticated Pacific white shrimp (*Penaeus vannamei*) line in China. *Sci. Rep.* 10 (1), 7840–7910. doi:10.1038/s41598-020-64597-x

Ren, S., Mather, P. B., Tang, B., and Hurwood, D. A. (2020c). Comparison of reproductive performance of domesticated *Litopenaeus vannamei* females reared in recirculating tanks and earthen ponds: An evaluation of reproductive quality of spawns in relation to female body size and spawning order. *Front. Mar. Sci.* 7, 560. doi:10.3389/fmars.2020.00560

Ren, S., Mather, P. B., Tang, B., and Hurwood, D. A. (2018). Levels of genetic diversity and inferred origins of *Penaeus vannamei* culture resources in China: Implications for the production of a broad synthetic base population for genetic improvement. *Aquaculture* 491, 221–231. doi:10.1016/j.aquaculture.2018.03.036

Ren, S., Mather, P. B., Tang, B., and Hurwood, D. A. (2022). Standardized microsatellite panels for pedigree management of farmed white-leg shrimp (*Penaeus vannamei*) stocks validated in a VIE tagged family selection line. *Aquaculture* 551, 737946. doi:10.1016/j.aquaculture.2022.737946

Ren, S., Prentis, P., Mather, P. B., Li, Y., Tang, B., and Hurwood, D. A. (2020a). Genetic parameters for growth and survival traits in a base population of Pacific white shrimp (*Litopenaeus vannamei*) developed from domesticated strains in China. *Aquaculture* 523, 735148. doi:10.1016/j.aquaculture.2020.735148

Robinson, N. A., Robledo, D., Sveen, L., Daniels, R. R., Krasnov, A., Coates, A., et al. (2022). Applying genetic technologies to combat infectious diseases in aquaculture. *Rev. Aquac.* doi:10.1111/raq.12733

Sasaki, O., Aihara, M., Nishiura, A., Takeda, H., and Satoh, M. (2015). Genetic analysis of the cumulative pseudo-survival rate during lactation of Holstein cattle in Japan by using random regression models. *J. Dairy Sci.* 98 (8), 5781–5795. doi:10.3168/jds.2014-9152

Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86 (1-3), 35–45. doi:10.1016/s0301-6226(03)00151-9

Schlicht, K., Krattenmacher, N., Lugert, V., Schulz, C., Thaller, G., and Tetens, J. (2018). Genetic analysis of production traits in turbot (*Scophthalmus maximus*) using random regression models based on molecular relatedness. *J. Anim. Breed. Genet.* 135 (4), 275–285. doi:10.1111/jbg.12337

Sewalem, A., Miglior, F., and Kistemaker, G. J. (2010). Analysis of the relationship between workability traits and functional longevity in Canadian dairy breeds. *J. Dairy Sci.* 93 (9), 4359–4365. doi:10.3168/jds.2009-2969

Suebsong, W., Poompuang, S., Srisapoome, P., Koonawootrittriron, S., Luengnaruemitchai, A., Johansen, H., et al. (2019). Selection response for *Streptococcus agalactiae* resistance in Nile tilapia *Oreochromis niloticus*. *J. Fish. Dis.* 42 (11), 1553–1562. doi:10.1111/jfd.13074

Tarrés, J., Bidanel, J. P., Hofer, A., and Ducrocq, V. (2006). Analysis of longevity and exterior traits on Large White sows in Switzerland. *J. Anim. Sci.* 84 (11), 2914–2924. doi:10.2527/jas.2005-707

Troell, M., Naylor, R. L., Metian, M., Beveridge, M., Tyedmers, P. H., Folke, C., et al. (2014). Does aquaculture add resilience to the global food system? *Proc. Natl. Acad. Sci. U. S. A.* 111 (37), 13257–13263. doi:10.1073/pnas.1404067111

Turra, E. M., de Oliveira, D. A. A., Valente, B. D., de Alencar Teixeira, E., de Assis Prado, S., de Melo, D. C., et al. (2012). Estimation of genetic parameters for body weights of Nile tilapia *Oreochromis niloticus* using random regression models. *Aquaculture* 354, 31–37. doi:10.1016/j.aquaculture.2012.04.035

Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (2017). Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet. Sel. Evol.* 49 (1), 17–13. doi:10.1186/s12711-017-0293-6

Van Pelt, M. L., Ducrocq, V., De Jong, G., Calus, M. P. L., and Veerkamp, R. F. (2016). Genetic changes of survival traits over the past 25 yr in Dutch dairy cattle. *J. Dairy Sci.* 99 (12), 9810–9819. doi:10.3168/jds.2016-11249

Van Pelt, M. L., Meuwissen, T. H. E., De Jong, G., and Veerkamp, R. F. (2015). Genetic analysis of longevity in Dutch dairy cattle using random regression. *J. Dairy Sci.* 98 (6), 4117–4130. doi:10.3168/jds.2014-9090

Van Sang, N., Luan, N. T., Van Hao, N., Van Nhien, T., Vu, N. T., and Nguyen, N. H. (2020). Genotype by environment interaction for survival and harvest body weight between recirculating tank system and pond culture in *Penaeus monodon*. *Aquaculture* 525, 735278. doi:10.1016/j.aquaculture.2020.735278

Vazquez, A. I., Gianola, D., Bates, D., Weigel, K. A., and Heringstad, B. (2009). Assessment of Poisson, logit, and linear models for genetic analysis of clinical mastitis in Norwegian Red cows. *J. Dairy Sci.* 92 (2), 739–748. doi:10.3168/jds.2008-1325

Veerkamp, R. F., Engel, B., and Brotherstone, S. (2002). Single and multitrait estimates of breeding values for survival using sire and animal models. *Anim. Sci.* 75 (1), 15–24. doi:10.1017/s1357729800052784

Vehviläinen, H., Kause, A., Koskinen, H., and Paananen, T. (2010). Genetic architecture of rainbow trout survival from egg to adult. *Genet. Res.* 92 (1), 1–11. doi:10.1017/S0016672310000017

Vu, N. T., Phuc, T. H., Oanh, K. T. P., Sang, N. V., Trang, T. T., and Nguyen, N. H. (2022). Accuracies of genomic predictions for disease resistance of striped catfish to *Edwardsiella ictaluri* using artificial intelligence algorithms. *G3-Genes Genom. Genet.* 12 (1), jkab361. doi:10.1093/g3journal/jkab361

Yáñez, J. M., Houston, R. D., and Newman, S. (2014). Genetics and genomics of disease resistance in salmonid species. *Front. Genet.* 5, 415. doi:10.3389/fgene.2014.00415

Yano, I. (2000). Cultivation of broodstock in closed recirculating system in specific pathogen free (SPF) penaeid shrimp. *Aquac. Sci.* 48 (2), 249–257.

Zavadilová, L., Němcová, E., and Štípková, M. (2011). Effect of type traits on functional longevity of Czech Holstein cows estimated from a Cox proportional hazards model. *J. Dairy Sci.* 94 (8), 4090–4099. doi:10.3168/jds.2010-3684

Fraslin, C., Koskinen, H., Nousianen, A., Houston, R. D., and Kause, A. (2022). Genome-wide association and genomic prediction of resistance to Flavobacterium columnare in a farmed rainbow trout population. *Aquaculture* 738332.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 18 October 2022
DOI 10.3389/fgene.2022.1007123



OPEN ACCESS

EDITED BY
Yuzine Esa,
Putra Malaysia University, Malaysia

REVIEWED BY
Song Jiang,
Chinese Academy of Fishery Sciences (CAFS), China
Jianghua Huang,
Chinese Academy of Fishery Sciences (CAFS), China

*CORRESPONDENCE
Md. Mehedi Hasan,
md.m.hasan@sydney.edu.au

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

RECEIVED 29 July 2022
ACCEPTED 27 September 2022
PUBLISHED 18 October 2022

CITATION
Hasan MM, Thomson PC, Raadsma HW and Khatkar MS (2022), Genetic analysis of digital image derived morphometric traits of black tiger shrimp (Penaeus monodon) by incorporating G × E investigations.
*Front. Genet.* 13:1007123.
doi: 10.3389/fgene.2022.1007123



# Genetic analysis of digital image derived morphometric traits of black tiger shrimp (*Penaeus monodon*) by incorporating G × E investigations

Md. Mehedi Hasan[1,2]*, Peter C. Thomson[1,2], Herman W. Raadsma[1,2] and Mehar S. Khatkar[1,2]

[1]The University of Sydney, Faculty of Science, Sydney School of Veterinary Science, Camden, NSW, Australia , [2]ARC Research Hub for Advanced Prawn Breeding, Townsville, QLD, Australia

The black tiger shrimp, *Penaeus monodon*, is the second most economically important aquaculture shrimp species in the world, and in Australia it is one of the most commonly farmed shrimp species. Despite its economic significance, very few studies have reported the genetic evaluation of economically important morphological size and shape traits of shrimp grown in commercial grow-out environments. In this study we obtained genetic parameter estimates and evaluated genotype-by-environment interaction (GxE) for nine body morphological traits of shrimp derived from images. The data set contained image and body weight (BW) records of 5,308 shrimp, from 64 sires and 54 dams, reared in eight grow-out ponds for an average of 133 days. From the images, landmark based morphological distances were computed from which novel morphological traits associated with size and shape were derived for genetic evaluation. These traits included body weight (BW), body length (BL), body size (BS), head size (HS), Abdominal size (AS), abdominal percentage (AP), tail tip (TT), front by back ratio (FBR), condition factor (CF) and condition factor length (CFL). We also evaluated G×E interaction effects of these traits for shrimp reared in different ponds. The heritability estimates for growth related morphological and body weight traits were moderately high (BW: $h^2$ = *0*.32 ± 0.05; BL: $h^2$ = *0*.36 ± 0.06; BS: $h^2$ = *0*.32 ± 0.05; HS: $h^2$ = *0*.31 ± 0.05; AS: $h^2$ = *0*.32 ± 0.05; and TT: $h^2$ = *0*.28 ± 0.05) and low for abdominal percentage and body shape traits (AP: $h^2$ = *0*.09 ± 0.02; FBR: $h^2$ = *0*.08 ± 0.02; CF: $h^2$ = *0*.06 ± 0.02; and CFL: $h^2$ = *0*.003 ± 0.004). G × E interaction were negligible for all traits for shrimp reared in different ponds, suggesting re-ranking is not prevalent for this population. Genetic correlations among growth related morphological traits were high ranging from 0.36 to 0.99, suggesting these traits can be simultaneously improved through indirect genetic selection. However, negative genetic correlations were observed for FBR & CF shape traits with major growth traits (ranged −0.08 to −0.95), suggesting genetic selection for rapid growth will likely result in thick/fatty shrimp over generations. Our study showed image-based landmark data can be successfully employed for genetic evaluation of complex morphological traits of shrimp and is potentially

amenable to machine-learning derived parameters in semi-automated high volume phenotyping systems needed under commercial conditions.

KEYWORDS

black tiger shrimp, breeding, morphological size, shape, heritability, genetic correlation, genotype-by-environment interaction

## Introduction

The black tiger shrimp (*Penaeus monodon*), hereafter shrimp, is one of the most commercially important aquaculture species in the world (Van Sang et al., 2020), and is the most popular cultivated shrimp species in Australia (APFA, 2016). However, most of the local demand is met by imports from overseas (FRDC, 2020). Due to this economic burden, and the significant biosecurity risks to Australia from imported products, there is a need for improvement of shrimp aquaculture in Australia. Traditionally, Australian shrimp aquaculture is dependent on the collection of wild broodstock for generating seedstock for stocking in aquaculture farms. However, a comparative study between domesticated and wild Australian shrimp populations revealed a 39% higher yield in the selectively bred stocks, suggesting a great potential of genetic improvement of this species in Australia (Norman-López et al., 2016). To address this, the application of genetic selection in shrimp will increase the genetic merit of locally bred stocks and farm profitability.

The most economically important traits in most shrimp breeding programs are body size and body weight, which are a measure of growth, and shape of the animal. Most aquaculture selective breeding programs focus on the genetic improvement of growth traits (e.g., weight and size), as it ensures higher yields. However, shape traits (e.g., key body conformation ratios, head size, abdominal size etc.) are equally important as candidate traits for genetic improvement (Cardoso et al., 2021). Shrimp that have more depth and width tend to yield more edible meat, compared to thinner and slender shaped ones. Moreover, animals which are uniform in size and shape are preferred by both industry and consumers. However, compared to finfish species, and due to difficulties in measuring shape and body shape index traits in shrimp, genetic evaluation of morphological traits in shrimp has not been performed. For example, due to natural curvature seen in the body structure of shrimp, it is difficult to compute body shape indexes, such as the condition factors. It has been demonstrated that landmark-based morphometric distance data derived from shrimp digital images can be used to evaluate and define complex body size and shape traits, such as body size and various aspects of morphological ratios (Hasan et al., 2020). This approach is robust and simple to assess complex traits, which are generally difficult to measure for shrimp, and therefore creates a novel opportunity to evaluate genetic parameters of these traits in shrimp.

An important aspect of all aquaculture production is to understand the underlying biological processes involved in size and shape of the animal, as it is reflected in the final product quality found in the marketplace. A key component of this is the estimation of genetic parameters associated with size and shape traits. Genetic improvement in shrimp aquaculture production can be made by developing selection programs utilizing these genetic parameter estimates. In Australia, genetic parameter estimates for growth traits have been previously studied for the black tiger shrimp in the tank environments (Kenway et al., 2006; Coman et al., 2010; Sun et al., 2015). However, no studies have been carried out so far in the pond environment, which is the usual commercial grow-out environment for most of the aquaculture production farms. It is very important to know the heritability estimates in pond environments rather than in the tanks, otherwise the estimates could be biased or confounded.

Furthermore, as with other aquaculture breeding programs, shrimp are often selected in one breeding station and are then distributed across different grow-out ponds or environments. To exploit the full genetic potential in the breeding program of a species, it is important to understand the performance and extent of realized genetic gains across environments. Genotype-by-environment interaction (G × E), which is the variable expression of phenotypes of genetically identical organisms in different environments, can occur if genotypes are not well adapted in their respective rearing environments. G × E reduces the efficiency of selection breeding programs, since the best-performing genotypes in one environment are not necessarily the best in another environment. Meta-analysis studies with both fish and shrimp species have reported the presence of significant G × E for growth traits (Sae-Lim et al., 2016; Hasan et al., 2020).

The present study was conducted to provide genetic parameter estimates of novel digital image-derived size and shape traits of shrimp from an Australian population of this species. In addition to this, we investigated the extent of G × E interactions in the shrimp population studied, which were raised across different grow-out ponds.

## Material and methods

### Family production, grow-out, pedigree construction and phenotype acquisition

All progeny in this experiment were sampled from commercial cohorts of *P. monodon* raised by Seafarms Group

Ltd., as described in Foote et al. (2019). Briefly, wild broodstock were sourced from Joseph Bonaparte Gulf, Northern Territory, Australia and transferred to a commercial hatchery at Flying Fish Point, Queensland, Australia. Broodstock maturation was conducted within indoor flow-through tank systems (density of 3 $m^{-2}$ at 28 ± 0.5°C) and broodstock were fed a commercial maturation diet. For each cohort, broodstock were allowed to mate naturally within the tanks, with any unmated females then artificially inseminated following industry practices. Females were spawned in communal spawning tanks and spawned eggs were transferred hatching tanks, and hatched nauplii were then transferred into 20,000-L larval rearing tanks (LRTs) at a density from 100 to 125 nauplii/L, and reared on a commercial diet (Ridley Aqua Feed, Australia) until 30 days of culture (DOC). LTRs were then pooled and stocked into seven 4000-$m^2$ grow-out earthen ponds and reared under commercial conditions at a density of 45 $m^{-2}$ until harvest. Preharvest ponds were immediately sampled by random cast net. From post-larval stage 15 (PL15) to harvest, the growth periods ranged from 124 to 143 days across ponds. In total 76 full sib and half-sib families were produced across 5,308 progeny and stocked across eight ponds as shown in Supplementary Material S1 and was similar to the population described by Noble et al. (2020b). Through the grow-out period, the key water quality parameters were recorded, including dissolved oxygen, temperature, pH and salinity.

For the genetic evaluation, 5,308 shrimp that were weighed and photographed, were included. Twelve landmarks' data points were manually captured from the photographed images of shrimp as described in Hasan et al. (2019). These consist of one anterior, two posterior, four dorsal and five ventral landmarks. From these landmark data points, 66 morphological distance measurements were derived from the Euclidean distances between all pairwise coordinates. A principal component analysis (PCA) using the "prcomp" function in R was performed to describe the pattern of shrimp shape variation using these 66 morphological distances. From the component loadings of the PCA, key traits associated with shrimp morphology were defined as 1) body length (BL), from the most anterior point of the antennal scale to most posterior point of the tail based on the first principal component (PC1), 2) front to back ratio (FBR) between front body area (head and first abdominal segment) and the back area (abdominal segment two, three and four) based on PC2, and 3) condition factor (CF) calculated as $(BW/BL^3) \times 100$, where BW = body weight, BL = body length, based partly on PC3. Moreover, additional important morphological traits were obtained from the landmark-derived morphological distance data, e.g. body size (BS), head size (HS) and abdominal size (AS) by summing appropriate triangular areas using Heron's formula. In addition, abdominal percentage (AP) was calculated as (AS/BS) × 100, tail tip (TT), the dorso-ventral distance across the tail, and condition factor length (CFL) calculated as $(BS/BL^2) \times 100$; where BS = body size, BL = body length.

Since tracing of broodstock contribution could not be done on farm, all potential broodstock were genotyped and parentage analysis was utilized to determine the contributing parents retrospectively as detailed by Guppy et al. (2020) and Noble et al. (2020a). A genotype-by-sequencing (GBS)-based approach using DArTSeq (Sansaloni et al., 2011) for genotyping of single nucleotide polymorphisms (SNPs) was used for the broodstock. This DArTSeq data set was used to derive a targeted 4 K DArTcap custom SNP panel of 4,194 SNPs for genotyping of the offspring (Guppy et al., 2020). CERVUS version 3.0.7 (Kalinowski et al., 2007) was used to perform family assignment, and Colony V2.0.6.4 (Jones and Wang, 2010) was employed to cluster the offspring to the genetic group when the parental information was missing, in which case an arbitrary parental ID was given to each group.

The family diversity of shrimp within each pond, which is a measure of the evenness of the distribution of families within ponds, was estimated by calculating the Shannon-Wiener Index using the "vegan" package in "R" (Oksanen et al., 2020). The index is calculated as

$$H = -\sum_{i=1}^{n} p_i \log_e p_i$$

where $p_i$ is the proportion of animals in the pond from family $i$; maximum diversity occurs when families are represented equally (all $p_i$ equal) resulting in $H_{max} = -\log_e n$. The proportional family overlaps between pairs of ponds were calculated by the Morisita-Horn index, employed in the "divo" package in "R" (Sadee et al., 2019). The index for a pair of ponds (1,2) is calculated as

$$MH = 2\sum_{i=1}^{n} p_{i1}\, p_{i2} \Big/ \left( \sum_{i=1}^{n} p_{i1}^2 + \sum_{i=1}^{n} p_{i2}^2 \right)$$

where $p_{i1}$ and $p_{i2}$ are the proportion of shrimp from family $i$, in ponds 1 and 2, and ranges from 0 (no overlap: entirely different families) to 1 (identical family representation).

## Statistical analysis

Initial exploratory analysis of the data was carried out by inspecting the distribution and homogeneity of variance assessed by performing Shapiro-Wilk and Levene's tests. An ANOVA was performed to evaluate the significance of fixed effect of ponds. Quantitative genetic analysis such as (co)variance components, heritabilities ($h^2$), and genetic correlations ($r_g$) between traits were estimated by restricted maximum-likelihood (REML) methods in ASReml-R 4.0 (VSNi) (Butler et al., 2017) by fitting the following animal model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of observations of each trait (namely, BW, BL, BS, HS, AS, AP, TT, CF, CFL, TT, and FBR), $\boldsymbol{\beta}$ is the vector of

TABLE 1 Shannon-Wiener diversity index of prawn families across ponds.

| Pond | 149 | 150 | 152 | 155 | 156 | 157 | 160 | 161 |
|---|---|---|---|---|---|---|---|---|
| Sire | 3.35 | 3.39 | 2.42 | 3.49 | 3.23 | 2.45 | 2.33 | 2.25 |
| Dam | 3.06 | 3.10 | 2.32 | 3.17 | 3.00 | 2.28 | 2.20 | 2.15 |

fixed effects ("pond", eight levels), **a** is the vector of the random animal additive genetic effects, and **e** is the vector of random residual effects. Further, **X** and **Z** are corresponding incidence matrices for **β** and **a**, respectively.

Heritability was estimated from the following equation: $h^2 = \sigma_A^2/(\sigma_A^2 + \sigma_e^2)$, where $\sigma_A^2$ = variance due to additive genetic components and $\sigma_e^2$ = variance arising from residual effects, respectively. A series of bivariate models were fitted to estimate covariance components among different traits in order to estimate genetic ($r_g$) phenotypic ($r_p$) and environmental ($r_e$) correlations. Bivariate animal models were also used to estimate genotype by environment interaction (G×E). Here, G×E was estimated by calculation of covariance components of the same trait of shrimp in different ponds (Falconer, 1952). Genetic correlations ($r_g$) were calculated from the following equation: $r_g = \sigma_{A12}/\sqrt{\sigma_{A1}^2 \times \sigma_{A2}^2}$, where $\sigma_{A1}^2$and $\sigma_{A2}^2$ are the estimated additive genetic variance components of the same trait in different ponds (labelled 1, 2) and $\sigma_{A12}$ is the estimated genetic covariance between the pair of ponds.

Indirect genetic selection was calculated as a correlated response in trait $y$ with 1 SD selection differential in trait $x$ from the following equation:

$$CR_y = r_g \times h_x \times h_y \times SD_y$$

where $SD_y$ is the SD of trait $y$.

The correlated response in trait $y$ as a percentage of gain possible from direct selection for trait $x$ is calculated as %IS, the relative efficiency of correlated response in trait $y$ when selection is applied on trait $x$ as a percentage of gain possible from direct selection for trait $y$, i.e.,

$$\%IS = \frac{CR_y}{SD_y} \times 100 = \left(r_g \times h_x / h_y\right) \times 100.$$

# Results

## Family genetic diversity and descriptive statistics

The population studied was comprised of 5,308 individual $F_1$ shrimp produced from 64 sires and 54 dams. These individual shrimp were raised in seven commercial ponds.

TABLE 2 Morisita-Horn overlap (similarity) index of prawn family distribution across ponds (a) sire (below diagonal) and (b) dam (above diagonal).

| Pond | 149 | 150 | 152 | 155 | 156 | 157 | 160 | 161 |
|---|---|---|---|---|---|---|---|---|
| **149** | | 0.96 | 0.56 | 0.90 | 0.96 | 0.52 | 0.15 | 0.16 |
| **150** | 0.95 | | 0.54 | 0.94 | 0.94 | 0.51 | 0.16 | 0.16 |
| **152** | 0.56 | 0.53 | | 0.39 | 0.55 | 0.95 | 0.02 | 0.03 |
| **155** | 0.89 | 0.93 | 0.39 | | 0.87 | 0.37 | 0.19 | 0.18 |
| **156** | 0.95 | 0.92 | 0.55 | 0.84 | | 0.50 | 0.17 | 0.18 |
| **157** | 0.52 | 0.50 | 0.95 | 0.37 | 0.50 | | 0.18 | 0.17 |
| **160** | 0.10 | 0.10 | 0.02 | 0.19 | 0.11 | 0.17 | | 0.97 |
| **161** | 0.11 | 0.10 | 0.03 | 0.18 | 0.13 | 0.16 | 0.97 | |

The analysis of Shannon-Wiener diversity index revealed that families were not evenly distributed across ponds (Table 1). For example, a few families were over-represented in some ponds, particularly Pond 161 which had the lowest diversity for both sire families ($H = 2.25$) and dam families ($H = 2.15$), with similar low diversity in Pond 160. In contrast, Pond 155 had the most even distribution of families (sire families: $H = 3.49$; dam families: $H = 3.17$). Family distributions across ponds were also assessed by calculating the Morisita-Horn index (Table 2). It measures the family similarity across ponds and ranges from zero (no overlap) to one (perfect overlap). Here, Ponds 160 and 161 showed a very high Morisita-Horn overlap index between them ($MH = 0.97$), the same ponds with low family diversity, but low overlap with all other ponds suggesting unequal family representation across ponds. Supplementary Material S1 shows the distribution of the 76 families across the eight ponds.

Mean, standard deviation (SD) and coefficient of variation (CV) of the traits studied are shown in Table 3. Shrimp had an average weight of 14.5 g, average body length of 10.8 cm, and abdominal percentage of 63.8% (Table 3). Body weight (BW) was the most variable trait (CV of 27%) while Abdominal percentage (AP) was the least variable (2.3%).

## Heritability estimates

Heritability ($h^2$) estimates were moderate for BW (0.32 ± 0.05), BL (0.36 ± 0.06), BS (0.32 ± 0.05), HS (0.31 ± 0.05) and AS (0.32 ± 0.05) and were low for AP (0.09 ± 0.02), FBR (0.08 ± 0.02), CF (0.08 ± 0.02) and CFL (0.003 ± 0.004). With the exception of CFL, all estimates were significantly greater than zero based on their small standard errors (SE) (0.004–0.06). However, for CFL the $h^2$ estimate was near zero, indicating presence of minimal additive genetic variability in this trait for selection. To evaluate the extent of family performance across ponds, estimated breeding values (EBVs) of sires and dams were obtained for each pond, and an

TABLE 3 Mean, standard deviation (SD), coefficient of variation (CV %), additive genetic variance ($\sigma_A^2$), residual variance component ($\sigma_e^2$) and estimated heritability ($h^2$) of morphological traits of shrimp ($n$ = 5,308).

| Trait | Mean | SD | CV (%) | $\sigma_A^2$ | $\sigma_e^2$ | $h^2$ |
|---|---|---|---|---|---|---|
| Body weight (g) (BW) | 14.52 | 3.92 | 27.0 | 4.56 ± 0.92 | 9.38 ± 0.52 | 0.32 ± 0.05 |
| Body length (cm) (BL) | 10.82 | 1.01 | 9.3 | 0.35 ± 0.07 | 0.61 ± 0.03 | 0.36 ± 0.06 |
| Body size ($cm^2$) (BS) | 13.39 | 2.61 | 19.5 | 2.08 ± 0.42 | 4.27 ± 0.23 | 0.32 ± 0.05 |
| Head size ($cm^2$) (HS) | 4.83 | 0.92 | 19.2 | 0.25 ± 0.05 | 0.56 ± 0.02 | 0.31 ± 0.05 |
| Abdominal size ($cm^2$) (AS) | 8.56 | 1.71 | 20.0 | 0.87 ± 0.17 | 1.80 ± 0.10 | 0.32 ± 0.05 |
| Abdominal percentage (%) (AP) | 63.87 | 1.48 | 2.3 | 0.15 ± 0.03 | 1.48 ± 0.03 | 0.09 ± 0.02 |
| Tail tip (TT) | 1.17 | 0.13 | 11.9 | 0.004 ± 0.001 | 0.01 ± 0.0005 | 0.28 ± 0.05 |
| Front by back ratio (FBR) | 0.84 | 0.06 | 7.2 | 0.0001 ± 0.00005 | 0.002 ± 0.00005 | 0.08 ± 0.02 |
| Condition factor (CF) | 1.11 | 0.07 | 6.6 | 0.0003 ± 0.0001 | 0.005 ± 0.0001 | 0.06 ± 0.02 |
| Condition factor length (CFL) | 11.30 | 0.52 | 4.6 | 0.0008 ± 0.001 | 0.25 ± 0.005 | 0.003 ± 0.004 |

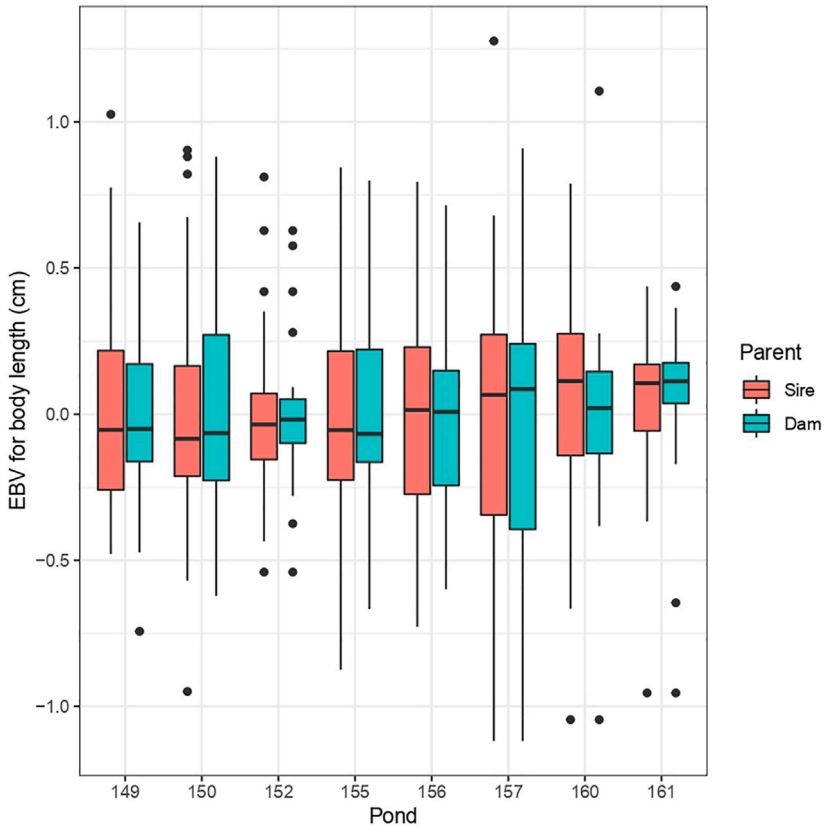



FIGURE 1
Distribution of EBVs for body length (BL) across ponds for sires and dams.

illustration of this is shown for body length in Figure 1. This considerable variability of EBVs is consistent with the estimated genetic SD for this trait (−0.6 cm), and this indicates there is potential for genetic improvement through selective breeding.

## Genetic correlations

Genetic correlations ($r_g$) among BW, BL, BS, HS, AS and TT were high (range: $r_g$ = 0.96–0.99) (Table 4). The $r_g$ was medium with AP trait with the abovementioned traits (range:

TABLE 4 Genetic ($r_g \pm$ se), phenotypic ($r_p \pm$ se), and environmental ($r_e \pm$ se) correlations among traits and extent of indirect selection response in correlated traits.

| Trait *x* | Trait *y* | $r_g \pm$ se | $r_p \pm$ se | $r_e \pm$ se | *$CR_y$ | % IS |
|---|---|---|---|---|---|---|
| BW# | BL | 0.99 ± 0.002 | 0.95 ± 0.001 | 0.94 ± 0.003 | 0.34 | 105 |
| BW | BS | 0.99 ± 0.001 | 0.95 ± 0.001 | 0.93 ± 0.003 | 0.32 | 99 |
| BW | HS | 0.99 ± 0.002 | 0.93 ± 0.002 | 0.90 ± 0.005 | 0.31 | 101 |
| BW | AS | 0.99 ± 0.002 | 0.93 ± 0.001 | 0.95 ± 0.003 | 0.32 | 99 |
| BW | AP | 0.36 ± 0.14 | 0.17 ± 0.01 | 0.13 ± 0.02 | 0.06 | 19 |
| BW | TT | 0.96 ± 0.01 | 0.86 ± 0.005 | 0.82 ± 0.009 | 0.29 | 103 |
| BW | FBR | −0.48 ± 0.13 | −0.19 ± 0.01 | −0.15 ± 0.02 | −0.08 | −96 |
| BW | CF | −0.63 ± 0.13 | 0.05 ± 0.01 | 0.16 ± 0.02 | −0.09 | −145 |
| BL | BS | 0.99 ± 0.0006 | 0.96 ± 0.001 | 0.95 ± 0.002 | 0.34 | 105 |
| BL | HS | 0.99 ± 0.001 | 0.94 ± 0.002 | 0.92 ± 0.004 | 0.33 | 107 |
| BL | AS | 0.99 ± 0.001 | 0.96 ± 0.001 | 0.95 ± 0.002 | 0.34 | 105 |
| BL | AP | 0.36 ± 0.14 | 0.15 ± 0.02 | 0.18 ± 0.01 | 0.06 | 72 |
| BL | TT | 0.97 ± 0.008 | 0.89 ± 0.004 | 0.85 ± 0.007 | 0.31 | 110 |
| BL | FBR | −0.49 ± 0.13 | −0.21 ± 0.01 | −0.16 ± 0.02 | −0.08 | −102 |
| BL | CF | −0.69 ± 0.12 | −0.17 ± 0.01 | 0.11 ± 0.02 | −0.1 | −169 |
| BS | HS | 0.99 ± 0.001 | 0.98 ± 0.0007 | 0.97 ± 0.001 | 0.32 | 99 |
| BS | AS | 0.99 ± 0.0003 | 0.99 ± 0.0002 | 0.99 ± 0.0004 | 0.32 | 99 |
| BS | AP | 0.40 ± 0.14 | 0.15 ± 0.01 | 0.10 ± 0.02 | 0.07 | 75 |
| BS | TT | 0.97 ± 0.007 | 0.92 ± 0.002 | 0.90 ± 0.004 | 0.29 | 104 |
| BS | FBR | −0.53 ± 0.12 | −0.17 ± 0.01 | −0.11 ± 0.02 | −0.08 | −106 |
| BS | CF | −0.67 ± 0.12 | −0.11 ± 0.01 | −0.03 ± 0.02 | −0.09 | −155 |
| HS | AS | 0.99 ± 0.003 | 0.95 ± 0.001 | 0.94 ± 0.003 | 0.31 | 97 |
| HS | TT | 0.97 ± 0.009 | 0.89 ± 0.004 | 0.85 ± 0.007 | 0.29 | 102 |
| HS | FBR | −0.45 ± 0.14 | −0.02 ± 0.01 | 0.06 ± 0.02 | −0.07 | −89 |
| HS | CF | −0.69 ± 0.12 | −0.11 ± 0.01 | −0.03 ± 0.02 | −0.09 | −157 |
| AS | TT | 0.97 ± 0.007 | 0.93 ± 0.002 | 0.91 ± 0.004 | 0.29 | 104 |
| AS | FBR | −0.56 ± 0.12 | −0.26 ± 0.01 | −0.21 ± 0.02 | −0.09 | −112 |
| AS | CF | −0.65 ± 0.12 | −0.11 ± 0.01 | 0.03 ± 0.02 | −0.09 | −150 |
| AP | FBR | −0.96 ± 0.01 | −0.88 ± 0.003 | −0.87 ± 0.003 | −0.16 | −51 |
| AP | CF | −0.08 ± 0.20 | 0.001 ± 0.01 | 0.008 ± 0.01 | −0.01 | −10 |
| TT | FBR | −0.57 ± 0.12 | −0.30 ± 0.01 | −0.26 ± 0.02 | −0.09 | −30 |
| TT | CF | −0.67 ± 0.12 | −0.12 ± 0.01 | −0.04 ± 0.02 | −0.09 | −31 |
| TT | CFL | 0.81 ± 0.31 | 0.40 ± 0.01 | 0.45 ± 0.01 | 0.02 | 8 |
| FBR | CF | 0.20 ± 0.20 | 0.003 ± 0.01 | −0.01 ± 0.01 | 0.01 | 23 |
| FBR | CFL | −0.96 ± 0.29 | 0.04 ± 0.01 | 0.08 ± 0.01 | −0.01 | −19 |
| CF | CFL | 0.28 ± 0.42 | 0.29 ± 0.03 | 0.29 ± 0.04 | 0.00 | 125 |

*Correlated response in trait *y* with 1 SD, selection differential in trait *x*. %IS, relative efficiency of correlated response in trait *y* when selection is applied on trait *x* as a percentage of gain possible from direct selection for trait *y*.

#BW, body weight (g); BL, body length; BS, body size; HS, head size; AS, abdominal size; AP, abdominal percentage; TT, tail tip; FBR, front-back ratio; CF, condition factor; CFL, condition factor length. Note that missing trait pairs correspond to when the bivariate model could not be fitted.

$r_g$ = 0.36–0.40), but highly negative with FBR ($r_g$ = -0.96). The phenotypic and environmental correlations were in similar alignment in both direction and magnitude as the genetic correlations for relationships between traits (Table 4).

## G × E across ponds

The genetic correlation between phenotypic expressions of the same traits in different ponds indicated lack of any G × E effect for all the traits studied (Figure 2, Supplementary Material S2). Although,

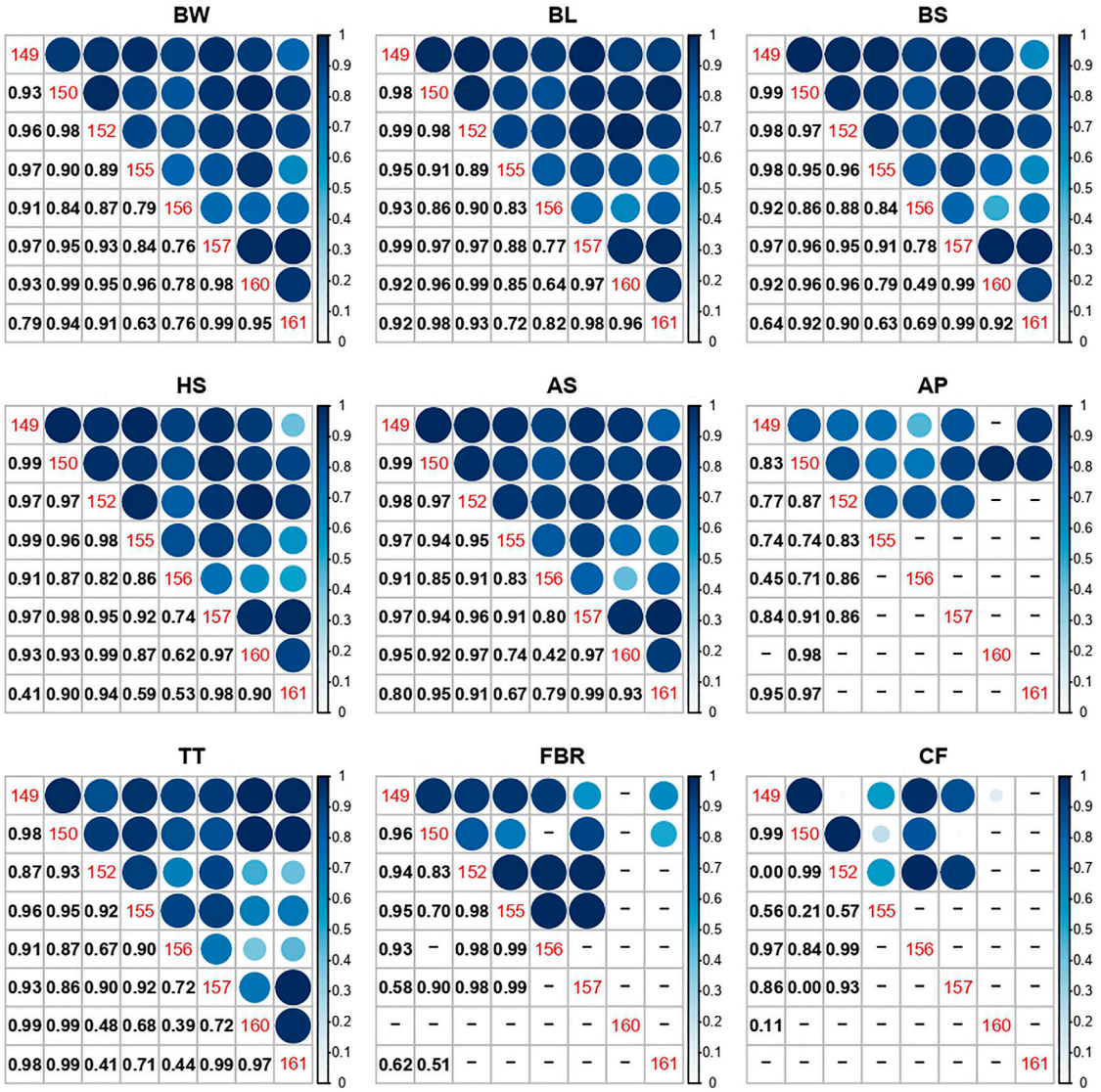



FIGURE 2
Correlation plots of estimated genetic correlations of traits between pairs of ponds, as a measure of G × E. Pond numbers are shown down the diagonal, with numerical values below the diagonal, and correlation visualisation above. Values shown as "–" are where the particular bivariate model could not be fitted. Trait codes are BW = body weight (g); BL = body length; BS = body size; HS = head size; AS = abdomenal size; AP = abdominal percentage; TT = tail tip; FBR = front-back ratio; CF = condition factor. Condition factor length (CFL) not shown as only one correlation could be estimated (Ponds 149 and 152: $r_g$ = 0.99).

there were indications of G × E among certain ponds (e.g., for BS trait between pond 155 & 161, 0.63 ± 0.45). However, the very high SE in these instances reflects the limited overlap of families in these pairs of ponds, limiting the ability to infer G × E or otherwise, for these ponds.

## Discussion

After *Penaeus vannamei*, *Penaeus monodon* is the next most important cultivated shrimp species in the world. However, compared to its aquaculture importance, genetic improvement of this species has been quite limited. Although some previous genetic parameter estimates have been reported for this species, they were all derived in tank environments. In commercial breeding programs it is important to simulate commercial conditions during genetic evaluation, so that the animals evaluated and selected for best targeted traits in the breeding nucleus population demonstrate similar yields in commercial settings. In this study, the performance of shrimp was evaluated in standard Australian farming conditions.

Growth traits are often considered the most important traits in any breeding programs due to their direct correlation with the economic value of the product. Overall, moderate to high heritability estimates were recorded for BW (0.32 ± 0.05), BL (0.36 ± 0.06) and BS (0.32 ± 0.05), suggesting substantial potential for genetic improvement of these traits in the *P. monodon* population studied. All these are growth-related traits and these findings corroborate the previous findings of *P. monodon* genetic parameter estimates (Kenway et al., 2006; Macbeth et al., 2007; Coman et al., 2010; Noble et al., 2020b), although a meta-analysis conducted by Hasan et al. (2020) reported the estimates ranged from 0.18 to 0.69. Similarly Noble et al. (2020b) reported the heritability estimate of 0.38 for BW trait of *P. monodon* for a smaller subset of this population. A higher heritability of 0.55 for growth rate (at 54 weeks) and 0.47 body weight were reported by Kenway et al. (2006) and Van Sang et al. (2020), respectively. However, Sun et al. (2015) reported low heritability estimates for body length (0.18) and body weight (0.24) trait of this species. Similarly, Krishna et al. (2011) also reported a low heritability estimate of 0.27 for body weight trait of *P. monodon*. There could be various factors that could lead to high level of heterogeneity of heritability estimates of any traits, including common family environmental effects (Rutten et al., 2005), characteristics of the grow-out farm (Turra et al., 2012), sex and age of the animal (Benzie et al., 1997) etc. For example, for body length, Benzie et al. (1997) reported a heritability of 0.59 at 6 weeks and 0.30 at 10 weeks of age for *P. monodon*.

In addition to these growth traits, heritability of traits related to shape e.g., HS (0.31 ± 0.05) and AS (0.32 ± 0.05) were moderately high, and low for AP (0.09 ± 0.02), FBR (0.08 ± 0.02), CF (0.06 ± 0.02) and CFL (0.003 ± 0.004) traits. This suggests that HS and AS traits are suitable for selection to change body shape of shrimp. Shrimp with smaller HS, larger AS and higher AP might be preferable for selection, as it may correspond directly to higher meat yield. Moreover, to attain uniformity of the cultured shrimp body shape, FBR trait can be selected in combination with these other traits by employing multi-trait selection criteria. Until now, no genetic analysis has been conducted on shape traits of shrimp. Quantitative trait loci (QTL) and genome wide association studies (GWAS) with Gilthead seabream (*Sparus aurata*), sea bass (*Dicentrarchus labrax*) and common carp (*Cyprinus carpio*) have shown that economically important shape-related traits are associated with unique genomic regions (Colihueque and Araneda, 2014). However, to date no specific study has been undertaken to search for genetic associations in shrimp for shape traits. Recently, due to increasing market sophistication, shape traits are gaining special attention by aquaculture breeders and consumers (Mehar et al., 2020). For example, high-backed and elliptical shaped common carp have high economic value and are commercially cultivated in fish farms. In the case of the ornamental goldfish (*Carassius auratus*), various morphological traits (e.g., body shape, fin morphology, and eye features) are modified *via* selective breeding to increase economic value (Colihueque and Araneda, 2014). Similarly, body shape traits of black tiger shrimp can be selected according to economic value and consumer preferences, e.g., shrimp with smaller heads may be preferred to increase meat yield, or shrimp with uniform size and shape may also be preferred by retailers and consumers (Mehar et al., 2020). This can be achieved through considering the genetic correlations of growth and morphological shape traits in the selective breeding program.

In our current study, genetic correlations between key growth-related morphological traits were positive and high (e.g., between and among BW, BL, BS, AS, and HS). This suggests that key traits for growth and shape of *P. monodon* can be selected indirectly. Our estimates of correlated responses among different traits indicates that selection on BS trait can indirectly improve BW and AP traits by 99% and 75%, respectively. This indicates these traits are a likely to be regulated by some of the same genes. A similar trend was observed among other morphological traits. Thus, selection of one such morphological trait will lead to change in other correlated morphological traits. In the case of rainbow trout, selection for increased BW indirectly resulted in fish with higher body width and height with rounded shape (Kause et al., 2003). However, a negative association was seen between FBR and CF traits with all other morphological traits, suggesting a potential trade-off is required in selecting shrimp body shape traits. This trade-off indicates the presence of certain limits in the selection of FBR (shape) trait, that is direct selection on these morphological traits will lead to decrease in FBR in this shrimp population, i.e., there will be thicker shaped shrimp. It is well known that functional interdependence among various traits plays a key role in the constraining the evolution of a certain shape. For example, there has been a report on the relationship between body shape and physical activity in fish (e.g., swimming performance) (Reid and Peichel, 2010). A similar underlying phenomenon may be responsible for the CF trade-off for shrimp, which warrants further investigation.

Genetic correlation estimates (as a measure of G×E) between ponds with high Morisita-Horn index scores (>0.56) were high for BW, BL, BS and HS. Only moderately low G×E ($r_g$ = 0.74 ± 0.29 and 0.45 ± 0.30) was observed in AP, for pond pairs of 149 and 155 and 149 and 156, respectively. This suggests that G × E interactions were non-existent for these traits and re-ranking of breeding values may be less pronounced across the ponds in this shrimp population, but this may be expected given all the ponds are in the one location. Similar to this finding, a meta-analysis, using data from 29 peer reviewed studies, found that growth-related morphological traits have a high genetic correlation ($r_g$ = 0.73 ± 0.05) across various environments and species, indicating low levels of G × E with low re-ranking of breeding values across environments (Hasan et al., 2020). Being a non-fitness trait, growth-related morphological traits are less likely to be affected by variation in specific environments (Mousseau and Roff, 1987; Visscher et al., 2008). However, G × E for growth-related traits is evident in shrimp species when the grow-out environment is stressful (e.g., temperature, salinity, ammonia tolerance etc.) (Coman et al., 2002; Li et al., 2015). This indicates re-ranking

can be present even for non-fitness traits (e.g., growth) when the environment is stressful.

For ponds with low family diversity (e.g., pond 160 and 161), low genetic correlation estimates were recorded for BW, BL, BS and HS. Although this could be an indication of high G × E, the estimates are unreliable due to their large standard error. This suggests the importance of maintaining homogenous family distribution across experimental grow-out environments, to obtain reliable G × E estimates.

For efficient improvement of traits under selection, the recording of these traits should be accurate and cost effective. Our study demonstrates that digital images can be used to derive economically-important growth-related morphological traits from landmark data to evaluate genetic merit. While in most breeding programs, body weight is the key trait for selection, it can be predicted from image-derived data as was the case in this study. So, if the breeding goal is to achieve a specific shape of the animal, and taking weight record is not feasible, then body measurements can be an excellent option to examine shrimp production traits. From our image-based analysis of morphological traits, we were able to derive size and shape measurements of the shrimp (e.g., length, width, abdominal percentage, abdominal size, FBR etc.) which can then be used for selection. For example, meat yield (peeled tail weight) of shrimp may not be feasible to record due to technical difficulties, however abdominal size can be used to infer the meat weight. Similarly, to select shrimp with a thick structure (FBR), body size (BS) trait can be selected. However, future studies are required for 1) phenotyping of morphological traits directly from the images using machine learning approaches without the need to do manual landmarking, and 2) obtaining estimates of economic values of the morphological traits, as these are unknown and need to be studied before incorporating in the breeding programs.

## Conclusion

Our study shows that landmark-based digital-image analysis is a promising tool of phenotyping of shrimp morphological traits and for genetic evaluation of these traits. Genetic improvement for growth-related morphological and weight traits is feasible, since these traits demonstrated high genetic variation and heritability. Family-level selection breeding is needed for genetic improvement of some shape traits (e.g., FBR), as the heritability was low. Within the same farm, it is not essential to perform separate genetic evaluations of shrimp across multiple ponds as G × E effects were found to be negligible between ponds.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

## Author contributions

All authors contributed to the manuscript. The study was conceived, designed and executed by MH, MK, HR, and PT. The manuscript was conceived and prepared by MH, MK, PT, and HR. PT addressed statistical and grammatical components of the manuscript. The final manuscript was reviewed by all authors, and all approved its contents.

## Funding

This work was in part provided by the Australian Research Council—Industrial Transformation Research Hub (IH130200013). MH was supported by a Research Training Program Stipend (RTP), Francis Henry Loxton Supplementary Scholarship from the Sydney Institute of Agriculture, and The Paulette Isabel Jones PhD Completion Scholarship, at The University of Sydney.

## Acknowledgments

We acknowledge the support we received from all the funding bodies and collaborators including all the members of the ARC Research Hub for Advanced Prawn Breeding, James Cook University, Australia. Access to the images was made available through the ARC Research Hub for Advanced Prawn Breeding, James Cook University, Australia as part of a PhD project to MH under supervision of MK, PT, and HR.

## Conflict of interest

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1007123/full#supplementary-material

# References


APFA (2016). Australian prawn farmer association: Australia in farmed prawns. AvaliableAt: https://apfa.com.au/australian-farmed-prawns/.

Benzie, J., Kenway, M., and Trott, L. (1997). Estimates for the heritability of size in juvenile *Penaeus monodon* prawns from half-sib matings. *Aquaculture* 152 (1-4), 49–53. doi:10.1016/s0044-8486(96)01528-1

Butler, D., Cullis, B. R., and Gogel, B. (2017). *ASReml-R reference manual version 4*. Hemel Hempstead, UK: VSN International Ltd.

Cardoso, A. J. d. S., Oliveira, C. A. L. d., Campos, E. C., Ribeiro, R. P., Assis, G. J. d. F., and e Silva, F. F. (2021). Estimation of genetic parameters for body areas in Nile tilapia measured by digital image analysis. *J. Animal Breed. Genet.* 138 (6), 731–738. doi:10.1111/jbg.12551

Colihueque, N., and Araneda, C. (2014). Appearance traits in fish farming: Progress from classical genetics to genomics, providing insight into current and potential genetic improvement. *Front. Genet.* 5, 251. doi:10.3389/fgene.2014.00251

Coman, G. J., Arnold, S. J., Wood, A. T., and Kube, P. D. (2010). Age: Age genetic correlations for weight of *Penaeus monodon* reared in broodstock tank systems. *Aquaculture* 307 (1-2), 1–5. doi:10.1016/j.aquaculture.2010.06.027

Coman, G. J., Crocos, P. J., Preston, N. P., and Fielder, D. J. A. (2002). The effects of temperature on the growth, survival and biomass of different families of juvenile *Penaeus japonicus* Bate. *Aquaculture* 214 (1-4), 185–199. doi:10.1016/s0044-8486(02)00361-7

Falconer, D. S. (1952). The problem of environment and selection. *Am. Nat.* 86 (830), 293–298. doi:10.1086/281736

Foote, A., Simma, D., Khatkar, M., Raadsma, H., Guppy, J., Coman, G., et al. (2019). Considerations for maintaining family diversity in commercially mass-spawned penaeid shrimp: A case study on *Penaeus monodon*. *Front. Genet.* 10, 1127. doi:10.3389/fgene.2019.01127

FRDC (2020). Fisheries research and development corporation: Seafood production and trade databases. AvaliableAt: https://www.frdc.com.au/seafood-production-and-trade-databases.

Guppy, J. L., Jones, D. B., Kjeldsen, S. R., Le Port, A., Khatkar, M. S., Wade, N. M., et al. (2020). Development and validation of a RAD-Seq target-capture based genotyping assay for routine application in advanced black tiger shrimp (*Penaeus monodon*) breeding programs. *BMC Genomics* 21 (1), 541–616. doi:10.1186/s12864-020-06960-w

Hasan, M. M., Tulloch, R. L., Thomson, P. C., Raadsma, H. W., and Khatkar, M. S. (2020). Meta-analysis of genetic parameters of production traits in cultured shrimp species. *Fish. Fish. (Oxf).* 21 (6), 1150–1174. doi:10.1111/faf.12495

Hasan, M. M., Wade, N. M., Bajema, C., Thomson, P. C., Jerry, D. J., Raadsma, H. W., et al. (2019). Evaluation of body morphology and shape of black tiger prawn (*Penaeus monodon*) by morphometric analysis. *Proc. Assoc. Adv. Animal Breed. Genet.* 23, 456–459.

Jones, O. R., and Wang, J. (2010). Colony: A program for parentage and sibship inference from multilocus genotype data. *Mol. Ecol. Resour.* 10 (3), 551–555. doi:10.1111/j.1755-0998.2009.02787.x

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16 (5), 1099–1106. doi:10.1111/j.1365-294X.2007.03089.x

Kause, A., Ritola, O., Paananen, T., Eskelinen, U., and Mäntysaari, E. (2003). Big and beautiful? Quantitative genetic parameters for appearance of large rainbow trout. *J. Fish Biol.* 62 (3), 610–622. doi:10.1046/j.1095-8649.2003.00051.x

Kenway, M., Macbeth, M., Salmon, M., McPhee, C., Benzie, J., Wilson, K., et al. (2006). Heritability and genetic correlations of growth and survival in black tiger prawn *Penaeus monodon* reared in tanks. *Aquaculture* 259 (1-4), 138–145. doi:10.1016/j.aquaculture.2006.05.042

Krishna, G., Gopikrishna, G., Gopal, C., Jahageerdar, S., Ravichandran, P., Kannappan, S., et al. (2011). Genetic parameters for growth and survival in *Penaeus monodon* cultured in India. *Aquaculture* 318 (1-2), 74–78. doi:10.1016/j.aquaculture.2011.04.028

Li, W., Luan, S., Luo, K., Sui, J., Xu, X., Tan, J., et al. (2015). Genetic parameters and genotype by environment interaction for cold tolerance, body weight and survival of the Pacific white shrimp *Penaeus vannamei* at different temperatures. *Aquaculture* 441, 8–15. doi:10.1016/j.aquaculture.2015.02.013

Macbeth, M., Kenway, M., Salmon, M., Benzie, J., Knibb, W., and Wilson, K. (2007). Heritability of reproductive traits and genetic correlations with growth in the black tiger prawn *Penaeus monodon* reared in tanks. *Aquaculture* 270 (1-4), 51–56. doi:10.1016/j.aquaculture.2007.03.018

Mehar, M., Mekkawy, W., McDougall, C., and Benzie, J. A. (2020). Fish trait preferences: A review of existing knowledge and implications for breeding programmes. *Rev. Aquac.* 12 (3), 1273–1296. doi:10.1111/raq.12382

Mousseau, T. A., and Roff, D. A. (1987). Natural selection and the heritability of fitness components. *Heredity* 59 (2), 181–197. doi:10.1038/hdy.1987.113

Noble, T. H., Coman, G. J., Wade, N. M., Thomson, P. C., Raadsma, H. W., Khatkar, M. S., et al. (2020a). Genetic parameters for tolerance to gill-associated virus under challenge-test conditions in the black tiger shrimp (*Penaeus monodon*). *Aquaculture* 516, 734428. doi:10.1016/j.aquaculture.2019.734428

Noble, T. H., Coman, G. J., Wade, N. M., Thomson, P. C., Raadsma, H. W., Khatkar, M. S., et al. (2020b). Genetic parameters of Gill-associated virus infection and body weight under commercial conditions in black tiger shrimp, *Penaeus monodon*. *Aquaculture* 528, 735580. doi:10.1016/j.aquaculture.2020.735580

Norman-López, A., Sellars, M. J., Pascoe, S., Coman, G. J., Murphy, B., Moore, N., et al. (2016). Productivity benefits of selectively breeding Black Tiger shrimp (*Penaeus monodon*) in Australia. *Aquac. Res.* 47 (10), 3287–3296. doi:10.1111/are.12782

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2020). Vegan: Community ecology package. *R Package Version 2*, 5–7.

Reid, D. T., and Peichel, C. L. (2010). Perspectives on the genetic architecture of divergence in body shape in sticklebacks. *Integr. Comp. Biol.* 50 (6), 1057–1066. doi:10.1093/icb/icq030

Rutten, M. J., Komen, H., and Bovenhuis, H. (2005). Longitudinal genetic analysis of Nile tilapia (*Oreochromis niloticus* L.) body weight using a random regression model. *Aquaculture* 246 (1-4), 101–113. doi:10.1016/j.aquaculture.2004.12.020

Sadee, C., Pietrzak, M., Seweryn, M., Wang, C., and Rempala, G. (2019). Diversity and overlap analysis package user's guide. Report.

Sae-Lim, P., Gjerde, B., Nielsen, H. M., Mulder, H., and Kause, A. (2016). A review of genotype-by-environment interaction and micro-environmental sensitivity in aquaculture species. *Rev. Aquac.* 8 (4), 369–393. doi:10.1111/raq.12098

Sansaloni, C., Petroli, C., Jaccoud, D., Carling, J., Detering, F., Grattapaglia, D., et al. (2011). Diversity arrays technology (DArT) and next-generation sequencing combined: Genome-wide, high throughput, highly informative genotyping for molecular breeding of *Eucalyptus*. *BMC Proc. Biomed. Cent.* 5, 1–2. doi:10.1186/1753-6561-5-S7-P54

Sun, M. M., Huang, J. H., Jiang, S. G., Yang, Q. B., Zhou, F. L., Zhu, C. Y., et al. (2015). Estimates of heritability and genetic correlations for growth-related traits in the tiger prawn *Penaeus monodon*. *Aquac. Res.* 46 (6), 1363–1368. doi:10.1111/are.12290

Turra, E. M., de Oliveira, D. A. A., Valente, B. D., de Alencar Teixeira, E., de Assis Prado, S., de Melo, D. C., et al. (2012). Estimation of genetic parameters for body weights of Nile tilapia *Oreochromis niloticus* using random regression models. *Aquaculture* 354, 31–37. doi:10.1016/j.aquaculture.2012.04.035

Van Sang, N., Luan, N. T., Van Hao, N., Van Nhien, T., Vu, N. T., and Nguyen, N. H. (2020). Genotype by environment interaction for survival and harvest body weight between recirculating tank system and pond culture in *Penaeus monodon*. *Aquaculture* 525, 735278. doi:10.1016/j.aquaculture.2020.735278

Visscher, P. M., Hill, W. G., and Wray, N. R. (2008). Heritability in the genomics era—Concepts and misconceptions. *Nat. Rev. Genet.* 9 (4), 255–266. doi:10.1038/nrg2322

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 08 November 2022
DOI 10.3389/fgene.2022.975174



OPEN ACCESS

EDITED BY
Yuzine Esa,
Putra Malaysia University, Malaysia

REVIEWED BY
Gonzalo Gajardo,
University of Los Lagos, Chile
Neil R. Loneragan Murdoch University, Australia

*CORRESPONDENCE
H. K. A. Premachandra,
ahoranek@usc.edu.au

†PRESENT ADDRESS
Wayne Knibb,
Aquabreed Pty Ltd., Diddillibah, QLD, Australia

‡These authors have contributed equally to this work and share first authorship

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

RECEIVED 22 June 2022
ACCEPTED 24 October 2022
PUBLISHED 08 November 2022

CITATION
Premachandra HKA, Becker A, Taylor MD and Knibb W (2022), Eastern king prawn Penaeus plebejus stock enhancement—Genetic evidence that hatchery bred prawns have survived in the wild after release.
*Front. Genet.* 13:975174.
doi: 10.3389/fgene.2022.975174



# Eastern king prawn *Penaeus plebejus* stock enhancement—Genetic evidence that hatchery bred prawns have survived in the wild after release

H. K. A. Premachandra[1*‡], Alistair Becker[2], Matthew D. Taylor[2] and Wayne Knibb[1†‡]

[1]Centre for Bioinnovation, University of the Sunshine Coast, Sippy Downs, QLD, Australia, [2]New South Wales Department of Primary Industries, Port Stephens Fisheries Institute, Nelson Bay, NSW, Australia

Eastern king prawn (*Penaeus plebejus*) is endemic to eastern Australia and is of high commercial and recreational value. As part of a recreational fisheries enhancement initiative, hatchery reared juveniles from Queensland were released into two, more Southern New South Wales (NSW) estuaries between 2014 and 2015. Responsible stock enhancement programs rely on knowledge of the population structure of the released species. Previously, in consideration of fisheries data, it was assumed the king prawn populations in Australia are one single breeding stock. In the present study, our first aim was to test this posit of no genetic differentiation using mtDNA control region (mtCR) sequences from the wild samples collected from four estuaries ranging from Queensland/NSW border (source of the stocked animals) to Southern NSW. The second objective was to test for signals of hatchery-released animals in the two stocked estuaries. All four surveyed populations had an extremely high level of haplotype diversity (average $h$ = 99.8%) and low level of haplotype sharing between populations. Estimates of PhiPT values were <0.01 or close to zero and AMOVA test did not indicate any significant differences among populations. Further, phylogenetic analysis and principal coordinate analysis did not support division of samples by population. Collectively these results suggest that eastern king prawn populations along the NSW coast can be considered as a single stock and stocking from the Queensland samples will not necessarily impact the genetic composition of the overall stock. After stocking of two estuaries, sharing of haplotypes was moderate to very high in the stocked sites (>80% in some collections) but negligible in the two unstocked estuaries (≤2%, which is assumed to be background coancestry unrelated to the hatchery). Moreover, some haplotypes present in the hatchery broodstock were detected in stocked sites, but not in unstocked sites. The highest stocking signal was detected in the estuary which becomes isolated from the sea by sand barrier suggesting such "lakes" maybe more favourable for stocking than estuaries directly open to the sea. Findings in the current study should assist in designing and implementation of future prawn stocking programs.

KEYWORDS

stock enhancement, genetic structure, *Penaeus plebejus*, eastern king prawn, mitochondrial control region

# 1 Introduction

Eastern king prawn *Penaeus plebejus* is a high value species in commercial and recreational fisheries and an endemic species to the eastern waters of Australia, ranging from a northern limit at Swain Reefs in Queensland down to Northeast Tasmania in the South (Ruello, 1975; Montgomery et al., 2007). The species has a complex life history involving an adult Northerly migration to spawning grounds in Northern New South Wales (NSW) and Southern Queensland, and Southward dispersal of spawned larvae along a Western boundary current (East Australian Current), followed by recruitment into estuarine nursery habitats (Montgomery, 1990). Here they form aggregations which are targeted by recreational anglers, establishing a socially and economically valuable niche recreational fishery (Reid and Montgomery, 2005; Taylor, 2017). However, many NSW estuaries are closed to the ocean for periods sometimes extending to years due to the longshore movement of sand (Roy et al., 2001). As a result, many estuaries are recruitment limited and these estuaries have been identified as potential locations for the fisheries enhancement programs involving the release of hatchery-reared post-larvae.

Aquaculture-based enhancement is a management approach primarily involving the release of cultured/hatchery-bred individuals into nursery habitats to enhance the existing wild stocks, and has been identified as a useful fisheries management tool over the past decades (Blankenship and Leber, 1995; Molony et al., 2003; Taiarui et al., 2019). These enhancements are usually either stock enhancement (where releases are intended to both enhance catch and contribute to spawning stock), sea ranching (where the primary objective of release is improved catch) or restocking (where stock rebuilding is the principal goal). Stock enhancement programs internationally often try to avoid deleterious genetic effects, such as a reduction of the genetic diversity of the wild stocks (which may reduce future adaptability) and also possible inbreeding of released siblings (which could lead to inbreeding depression) (e.g., Blankenship and Leber, 1995; Taylor et al., 2005; Lorenzen et al., 2010). Assessment of the genetic structure of the existing wild populations and the hatchery stocks is typically integral to responsible enhancement programs.

Prawn releases have occurred for over 50 years in various countries, at large scales in China and Japan, but also in Kuwait, United States, Taiwan, Australia and Sri Lanka (Davenport et al., 1999; Loneragan et al., 2006; Setio, 2016) resulting in varying levels of success (Kitada, 2018; Serajuddin et al., 2018; Kitada, 2020). Highly variable and low recapture rates were reported from several prawn stock enhancement programs in Japan where in the majority of cases stocked prawns represented 2% or less of the catch and only two cases exceeded 10% recapture rate (Hamasaki and Kitada, 2006). However, some other stock enhancement programs for *Penaeus japonicus* and *P. chinensis* reported high recovery rates (stocked animals were identified by tags and uropod clipping) varying from 4% to 35.6% (Liu, 1990; Loneragan et al., 2006; Wang et al., 2006). A pilot stock enhancement program conducted for eastern king prawn in South-Eastern Australia identified contribution levels up to 50%, however this contribution was calculated based on very small sample size ($n = 27$) (Setio, 2016). The high recapture rate may be attributed to the estuary (Lake Tyers) being closed to the ocean, preventing stocked prawns emigrating out the mouth (Setio, 2016) highlighting how stock enhancements can benefit fisheries in such situations.

Reliable methods for identification of hatchery individuals at recapture is crucial for the precise assessment of stock enhancement programs. For some prawn release programs, stocked animals have been identified using coded wire tags and uropod clipping, as in Japan (Hamasaki and Kitada, 2006), and some enhancement programs were assessed simply considering harvest yield changes/gains (Davenport et al., 1999). Genetic marker technologies have been used in many finfish stock enhancement programs, such as to estimate the contribution stocked fish in recreational or commercial catches (e.g., Taylor et al., 2021) and to trace the origin of fish where hatchery information is incomplete or not available (Hsu et al., 2020). Different types of genetic markers are being used for population and pedigree analyses, from DNA microsatellite markers to single nucleotide polymorphisms (SNPs) (Premachandra et al., 2017; Premachandra et al., 2019; Knibb et al., 2020), but provided there is substantial diversity, we have found the use of mitochondrial haplotype sequences is often cost effective and practical (McMillen-Jackson and Bert, 2004; Premachandra et al., 2017). Moreover, as mitochondrial DNA is matroclinously inherited, we can often infer family linages (Niwa et al., 2003; Cothran et al., 2005; Knibb et al., 2014b). To date, genetic detection technology is comparatively rare for detecting released crustaceans except a few cases reported for crabs and prawns (Obata et al., 2006; Setio, 2016; Wang et al., 2016; Liu et al., 2018).

A 2-year release program for eastern king prawn was initiated in recruitment limited estuaries in Australia in 2014, to enhance recreational fisheries for the species. The associated monitoring program collected prawn samples and assessed abundance from stocked and unstocked estuaries and this statistical information was reported to provide an initial non genetic determination of the potential contribution of the releases to prawn populations within the estuaries (Becker et al., 2018). The objectives of the present study were to, first, use the above collected samples, and


FIGURE 1
Geographical locations of eastern king prawn sample collected estuaries in New South Wales and Rocky Point hatchery in Queensland, Australia. Durras and Corunna estuaries were non-stocked reference systems, and Wallagoot and Tabourie estuaries were stocked with hatchery releases. Maps available in Wikimedia Commons web pages (https://commons.wikimedia.org/wiki/File:New_South_Wales_in_Australia.svg and NordNordWest/Wikipedia; "https://commons.wikimedia.org/wiki/File:Australia_New_South_Wales_location_map_blank.svg" \o "https://commons.wikimedia.org/wiki/File:Australia_New_South_Wales_location_map_blank.svg"https://commons.wikimedia.org/wiki/File:Australia_New_South_Wales_location_map_blank.svghttps://commons.wikimedia.org/wiki/File:Australia_New_South_Wales_location_map_blank.svg) were adapted under the Creative Commons license (https://creativecommons.org/licenses/by-sa/3.0/de/deed.en).

further samples, to assess the genetic structure of the wild eastern king prawn populations collected from four estuaries from Queensland/NSW border to Southern NSW before stock enhancement. Our second objective was to test for signals of the hatchery-released animals during post-release monitoring in the stocked estuaries using genetic measures, namely mitochondrial control region sequence.

# 2 Materials and method

## 2.1 Prawn stocking and sample collection

Wild broodstock consisted of female prawns that had already received spermatophores. They were collected by commercial trawlers and transported to hatchery facilities at Rocky Point (Figure 1). In the hatchery, females spawned naturally, or in some cases spawning was induced by eyestalk enucleation (Kelemec and Smith, 1980). Prawns were reared to post-larvae size (~18 mm TL) over 20 days before being transported to the release locations. Stocking occurred in two estuaries during December of 2014 and again in December of 2015 (See Table 1 for release locations and numbers) with greater numbers released during 2015.

Four estuaries (Figure 1) consisting of two non-stocked reference systems (Durras and Corunna) and two stocked systems (Tabourie and Wallagoot) were sampled nocturnally for prawns using a cast net "Fitech, Memphis, TN, USA, 2.45 m diameter, 4.75 mm square monofilament mesh". Eastern king prawn were collected during multiple sampling trips from all estuaries other than Wallagoot prior to the first 2014 stocking event (Supplementary Table S1). No prawns were sampled from Wallagoot as we understood that the estuary had not received a natural recruitment event in several years and none could be collected despite intensive sampling efforts. After the 2014 stocking, multiple sampling trips were made to each of the estuaries until October 2016 (Supplementary Table S1), aided by citizen-science sampling programs to collect prawns as they grew, resulting in between 168 and 253 samples per estuary (Table 1). Wallagoot estuary remained closed to the ocean after the 2014 stocking, meaning eastern king prawn samples collected before October 2015 (time when the estuary did open to the sea), were all suspected to be of hatchery origin. In addition to the prawns sampled from the estuaries, 46 wild broodstock were also collected from the Rocky Point hatchery facilities that were used for the 2015 stocking event (no broodstock were retained prior to the 2014 stocking). However, the 46 broodstock may not have represented the complete set of broodstock for the 2015 event and an unknown portion were lost. Tissue samples were placed into 2 ml vials containing 85% ethanol and stored at 4 °C until used for genomic DNA extraction.

## 2.2 DNA extraction and PCR amplification of mtCR

Genomic DNA extraction was conducted using QIAGEN DNeasy Blood and Tissue Kit, following manufacture's protocol. The integrity of extracted DNA was tested using 1% agarose gel electrophoresis. PCR amplification of mtDNA control region (mtCR) was achieved using a published primer pair (Chan et al., 2014). PCR assays were conducted in 30 µL reaction volumes containing, 0.2 µM of each primer (forward 5′-ATTAGCACTAGGTACTGAGA-3′and reverse 5′-AGTTTC AGGATAAGAAGACACTAT-3′) in 2 µL, 15 µL of 2 × PCR Master Mix, 11 µL of RNase free water, and 2 µL of the DNA template. Amplification was performed using an Eppendorf Mastercycler Nexus (Hamburg, Germany) with pre-denaturation at 94°C for 5 min; 35 cycles of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, extension at 72°C for 50 s; and final extension at 72°C for 10 min. Amplicons were

TABLE 1 The number of eastern king prawns stocked and sampled by year of released and estuary.

| Estuary | Treatment | 2014 stocking | 2015 stocking | Samples taken |
|---|---|---|---|---|
| Durras | References | 0 | 0 | 168 |
| Corunna | References | 0 | 0 | 170 |
| Tabourie | Stocked | 260,000 | 1,305,000 | 234 |
| Wallagoot | Stocked | 726,000 | 3,501,000 | 253 |

tested on 1.5% agarose gels stained with ethidium bromide to ensure successful amplification.

DNA sequencing of PCR products was conducted using DNA Sanger sequencing (AGRF, Australia) from both directions using the specific PCR primer pair mentioned above. Sequences were evaluated for the sequencing errors where there was disagreement between the forward and reverse sequences, the electropherograms were inspected manually, or when one of the two sequences corresponded to the consensus sequence, the consensus sequence was used.

## 2.3 Genetic diversity and population structure analysis

Genetic diversity within and among populations was estimated using mtCR sequences. Sequences were aligned (ClustalW) and tested for DNA sequence similarity within and between sample locations using BioEdit Sequence alignment editor V7.0.5.3 (Hall, 1999). Haplotype diversity, nucleotide diversity, number of polymorphic loci and presence of private alleles were tested using DnaSP v.5.0 (Librado and Rozas, 2009) and/or GenAIEx V6.5 (Peakall and Smouse, 2012). Genetic diversity among population were tested using PhiPT estimates, an $F_{ST}$ analogue which calculates population differentiation based on the genotypic variance (Teixeira et al., 2014) and analysis of molecular variance (AMOVA) among populations was conducted using GenAIEx V6.5 (Peakall and Smouse, 2012). Phylogenetic analysis for mtCR sequences was carried out implementing Maximum Likelihood method. Prior to the construction of phylogenetic tree, an assessment was conducted to find the best nucleotide substitution model and GTR + G + I (General Time Reversible + Gamma distribution + evolutionary invariable) model was selected as the best-fitting nucleotide substitution model for the maximum likelihood analysis based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores using MEGA 11 (Tamura et al., 2021). The pattern of genetic relationship among samples was created using principal coordinate analysis (PCoA) test available in GenAlEx V6.5 (Peakall and Smouse, 2012) based on the haploid genetic distance matrix to examine any population structure due to geographical proximity. The first two axes explaining the majority of the variance among the samples were used to obtain the PCoA plot.

# 3 Results

## 3.1 Genetic diversity and similarities among sites

### 3.1.1 Basic descriptions on mtCR haplotype and allelic variation

A stretch of 716 bp sequence fragment was recovered from 871 samples. Considering all samples, 294 polymorphic sites were detected along the 716 bp length, and a total of 750 mtCR haplotypes were identified among 871 samples and five populations (GenBank accessions ON804899—ON805769). The average haplotype diversity ($h$) and average nucleotide diversity ($\pi$) was 99.8% and 2.3%, respectively.

Considering all samples from the different times of sampling (all 871 samples collected across both years), each separate population had at least some specific types of haplotypes shared between individuals within that given population (Supplementary Figure S1A), with the notable exception of wild-collected broodstock sampled from Rocky Point Hatchery, where every sample had a different (i.e., unique), haplotype. When sharing occurred it was at very low level for Durras (of 167 different haplotypes recorded, only one was shared i.e., one haplotype was in more than one animal) and also low at Corunna (168 and two, respectively) and a little higher in Tabourie (227 and seven, respectively) and highest in Wallagoot (168 and nine, respectively). Considering sharing among different populations, most populations shared some haplotypes with each other, albeit at low levels (a minimum of 0 to a maximum of 6 haplotypes, Supplementary Figure S1B).

### 3.1.2 Statistical analyses comparing sites and times of sampling

There was a total of 44 different collections made over the four estuaries, at different times, both before and after stocking. Excluding the data from the stocked sites after stocking, pairwise PhiPT analyses of each sample at each time indicated there were 20 significant cases out of a total of 210 comparisons at $p < 0.05$ (Supplementary Table S2). Due to the relatively low level of

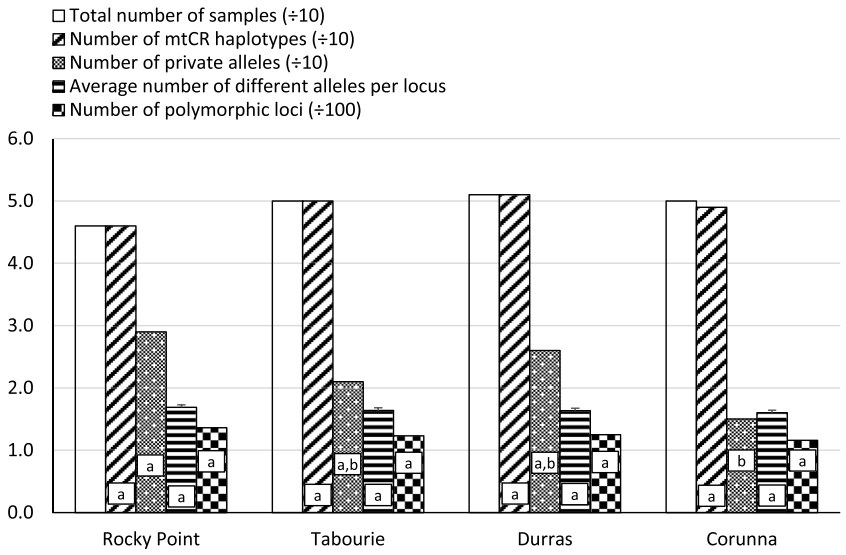



FIGURE 2
The number of mtCR haplotypes and allelic pattern distribution among four eastern king prawn populations. The number of mtCR haplotypes is the number of different/unique mtCR sequences found in a given population. The number of private alleles is the number of mtCR alleles (at a given nucleotide position) that occurred only in one population and not in others. The average number of different alleles is the number of different alleles at a given nucleotide loci, averaged over all nucleotide loci. The number of polymorphic loci is the number of nucleotide positions which were polymorphic (i.e., that had more than one nucleotide base). The analyses were conducted using samples collected before release from each site, and sample sizes were close to 50 for each site (Supplementary Table S1). No samples from Wallagoot were available before stocking. Bars with the same letters are not significantly different form each other ($p > 0.05$). For mtCR haplotype counts, number of private alleles and number of polymorphic loci, significance levels were determined using **(A)** Chi-square tests based on total counts and **(B)** using the Bonferroni method to correct for multiple comparisons. For average number of different alleles, significance levels were assessed using **(A)** ANOVA and **(B)** Bonferroni post hoc test for multiple comparisons.

**TABLE 2 Pairwise population differences measured using PhiPT values among four eastern king prawn populations using samples collected before stocking. PhiPT values are shown below diagonal, and the statical significance of these values are (probability, based on 999 permutations) shown above diagonal. No samples from Wallagoot were available before stocking. *n* = sample sizes.**

| | Rocky point | Tabourie | Durras | Corunna |
|---|---|---|---|---|
| Rocky point ($n$ = 46) | — | 0.397 | 0.448 | 0.414 |
| Tabourie ($n$ = 50) | 0.000 | — | 0.092 | 0.456 |
| Durras ($n$ = 51) | 0.000 | 0.008 | — | 0.229 |
| Corunna ($n$ = 50) | 0.000 | 0.000 | 0.003 | — |

significant cases (only a little above that expected by chance) and their temporally inconsistent nature, we have pooled samples from estuaries across the multiple sampling times prior to stocking. After pooling the temporal samples there were 195 haplotypes for the total of 197 samples. Statistical tests comparing between estuaries using the samples collected before release, indicated most measures of genetic diversity did not significantly vary among estuaries (Figure 2).

### 3.1.3 Tests of population genetic structure

Pairwise comparison of PhiPT values for samples collected before stocking did not indicate any significant genetic structure among populations (Table 2). All the PhiPT values were <0.01 or close to zero. AMOVA test also did not identify any significant differences among populations (PhiPT = 0.001; $p$ = 0.429) and all the variation (100%) was attributable to within population variation. DNA sequence divergence estimates were very similar for within populations and among populations comparisons, varying between 2.25%–2.29% (Table 3).

Phylogenetic (Maximum-Likelihood tree) analysis of the mtCR sequences generally did not support division of samples based on the populations (Figure 3) with some exceptions involving animals from Wallagoot (after stocking) where some

TABLE 3 Mean percent mtCR DNA sequence divergence for within and between four eastern king prawn populations, using samples collected before stocking. The percent divergence, for a given data point in the table, is calculated as the divergence in nucleotide sequence between pairs of samples, averaged over all pairs of samples. Mean percent divergence for within populations is shown in shaded cells along the diagonal. No samples from Wallagoot were available before stocking. *n* = sample sizes.

| | Rocky point | Tabourie | Durras | Corunna |
|---|---|---|---|---|
| Rocky point (*n* = 46) | 2.29 | — | — | — |
| Tabourie (*n* = 50) | 2.29 | 2.26 | — | — |
| Durras (*n* = 51) | 2.28 | 2.27 | 2.26 | — |
| Corunna (*n* = 50) | 2.29 | 2.25 | 2.25 | 2.28 |

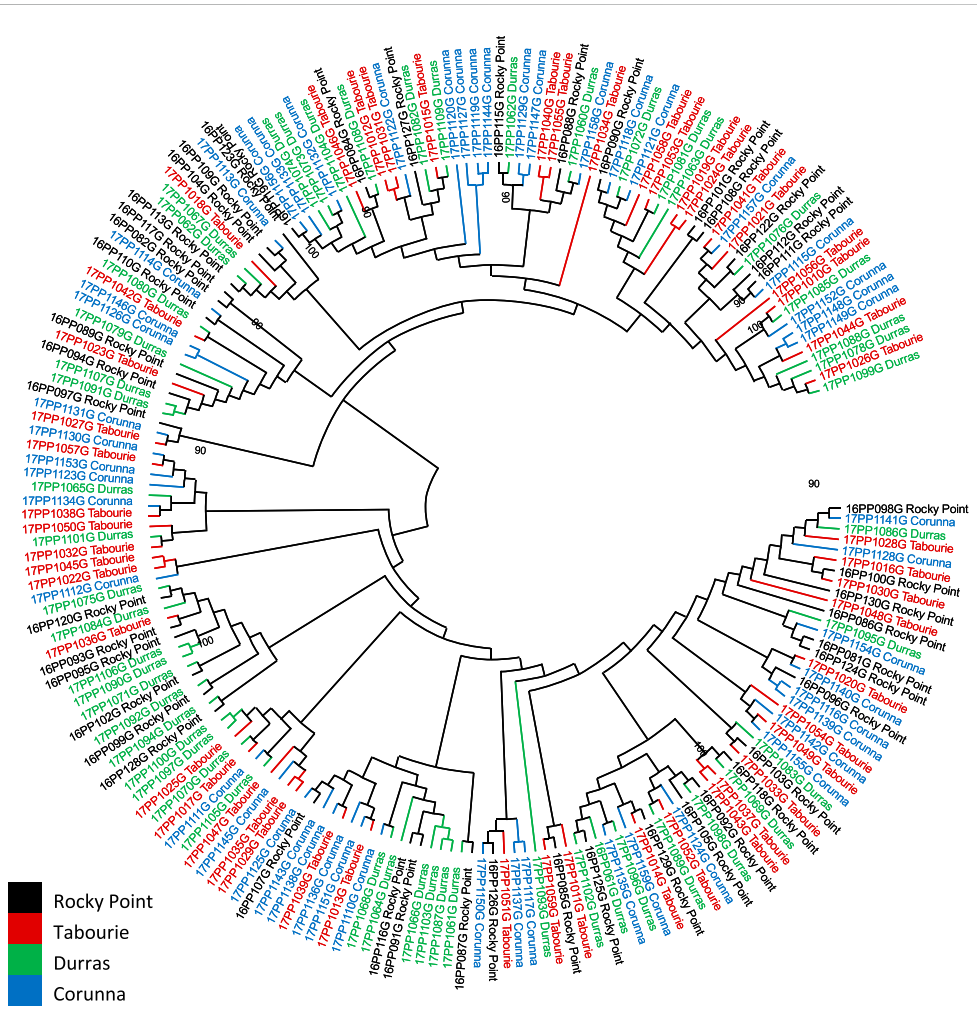



FIGURE 3
Phylogenetic analysis of the eastern king prawn mtCR sequences from four populations collected before stocking. No samples from Wallagoot estuary were available before stocking. Analysis was conducted using maximum likelihood method using GTR + G + I model in MEGA11 (Tamura et al., 2021). Each sample is colour coded based on the sample location, as indicated within the image. Only bootstrap values >90 are included as text in the figure (bootstrap value >90 offer statistical support for the grouping of the samples).

similar haplotypes were grouped together with high bootstrap support values (Supplementary Figure S2). PCoA analysis reveals no genetic structure among the populations from various locations and the genetic relatedness among the samples did not form any distinct cluster based on locations (Supplementary Figure S3).

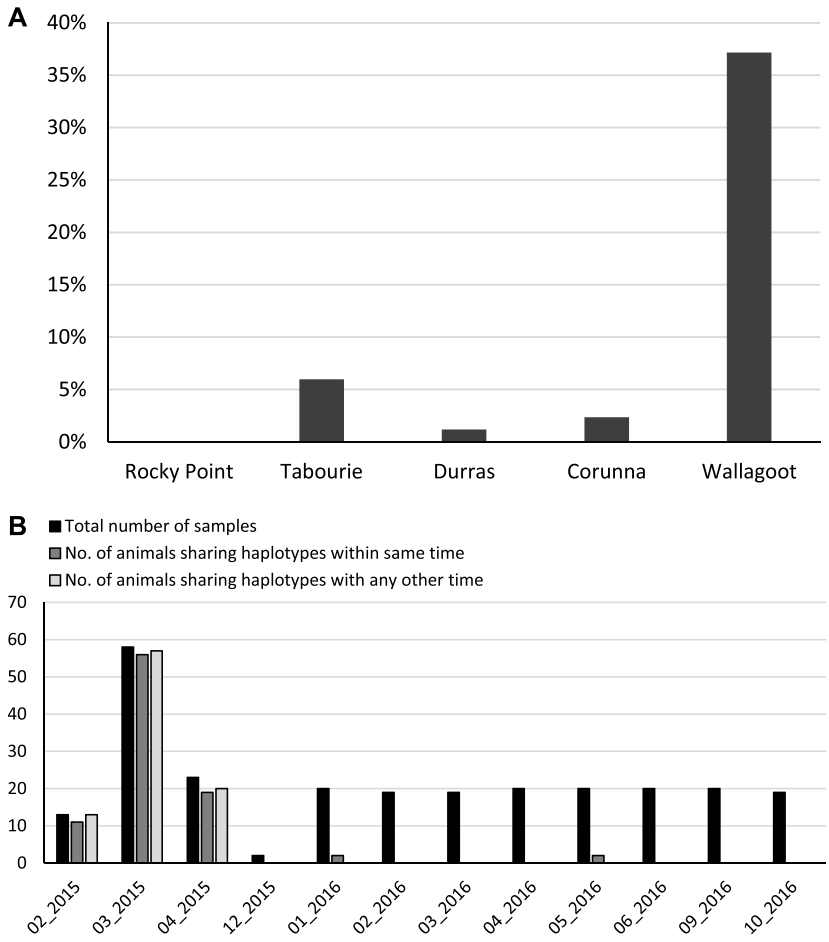



FIGURE 4
The level of mtCR haplotype sharing between animals within populations. **(A)** considering all populations and all samples, bars represent the percentage of animals that shared mtCR haplotypes within each population. **(B)** considering only the Wallagoot site, bars represent the number of animals sharing mtCR haplotypes by sample collection time.

## 3.2 Evidence for contribution of hatchery-releases

### 3.2.1 Haplotype sharing within sites

For the unstocked estuaries Corunna and Durras, nearly every animal had its own unique mtDNA haplotype, specifically 99% and 98% of haplotypes were unique respectively. By contrast, only 63% of the samples from Wallagoot, and 94% of the samples from Tabourie had unique mtDNA haplotypes (Figure 4A). A total of 94 samples out of the total of 253 samples shared haplotypes at Wallagoot, consisting of 9 shared haplotype groups. Specifically, over 80% of the samples collected within the first 5 months after the 2014 release in Wallagoot had shared haplotypes (Figure 4B).

Pairwise PhiPT values estimated considering the different sample collection time cohorts indicated that the early collections of Wallagoot samples within the first 5 months of stocking were significantly different from the later samples (Supplementary Table S2). Considering standard number of samples per estuary (i.e., $n$ = 46) selected randomly irrespective of the sample collection time from each site, Wallagoot samples had the lowest number of haplotypes among all the populations (Supplementary Figure S4), and the individuals sharing the same haplotypes were grouped together with high bootstrap support values in the phylogenetic analysis (Supplementary Figure S2).

### 3.2.2 Detection of "rocky point" hatchery haplotypes in other sites

As a further indication of a stocking signal, we tested the presence of matching haplotypes between hatchery and other populations. Considering all the samples, only the two stocked estuaries had some individuals with the same haplotypes as those found in the hatchery broodstock individuals but at low

frequencies (Supplementary Figure S5). Specifically, these matching haplotypes were found in samples collected after the second stocking event (when we had samples from some broodstock), except for one individual from Tabourie that was sampled before stocking.

# 4 Discussion

## 4.1 Genetic diversity

The observed haplotype diversity for eastern king prawn was very high, especially in the unstocked reference estuaries where more than 99% of animals had a unique haplotype. Previous work has shown similar patterns in diversity, with private mtCR haplotypes across an entire sample of banana prawn *Fenneropenaeus merguiensis* (Knibb et al., 2014b) and pink shrimp *Farfantepenaeus duorarum* (McMillen-Jackson and Bert, 2004). Various factors could contribute to this high level of variation, including the likelihood of very high effective population sizes of wild prawns, and also hypervariability of the mtCR, relative to coding genes (Stoneking, 2000; Diniz et al., 2005). Furthermore, the mtCR has an apparent high mutation rate of 19%/MY, as reported for the brown shrimp (McMillen-Jackson and Bert, 2003). Indeed, McMillen-Jackson and Bert (2004) suggested that mtCR in penaeids might be a useful genetic marker. Also, the high proportion of unique haplotypes may indicate the presence of large number of different maternal families within the spawning populations along the East coast of Australia. This assumption is supported by the recent stock assessment report for eastern king prawn, that indicated a large spawning biomass (Helidoniotis et al., 2020).

Despite this high level of haplotype diversity (which can be due to a single nucleotide change over about 700 bp), there was a high level of DNA similarity across the ~700 bp with an average of 98% similarity for all sample pairs. DNA sequence divergence between sites were very similar to the within site divergence. Consequently, most of the tests between sites (using pre-release samples and pooling the different temporal samples at each site) were not statistically significant. None of the pairwise comparison of PhiPT values were significant between sites for pooled temporal data, and 100% of the molecular variation was attributable to within population variation. Moreover, neither phylogenetic analysis, nor PCoA analysis identified segregation of samples by populations.

Collectively, these results suggest that the eastern king prawn populations in the east coast of Australia can be considered as a single panmictic genetic population, a conclusion which agrees with previous considerations based on adult migration, spawning patterns and dispersal of larvae along the coast (Montgomery, 1990; Montgomery et al., 2007; Taylor and Johnson, 2021), and also unpublished preliminary genetic work from a Ph. D. thesis (Chan, 2015). The combination of a Northward adult migration to spawning grounds (Ruello, 1975; Taylor and Johnson, 2021) and dispersal of larvae in Southwards flowing currents, likely explain the lack of any genetic structuring with the eastern king prawn population. Although there is likely to be annual variation in larval transport owing to changes in the conditions of the East Australian Current, modelling has shown the mean dispersal distances can range between ~750 and 1,000 km before settlement (Everett et al., 2017). This provides significant opportunity for mixing to occur during the southward phase of their life history, which would contribute to the lack of structure observed in this study.

## 4.2 Evidence of contribution of released prawns

We posit that, given the high level of haplotype diversity, that shared haplotypes within sites result from kinship (i.e., individuals being members of the same families). For Wallagoot, a stocked site, a very large number of samples (~37%), share a haplotype with at least one other sample. Indeed, within about the 6 months of the first stocking for Wallagoot, when the estuary was closed to the ocean, there is genetic evidence that nearly all samples were from stocked animals, based on the criteria of sibship. Approximately 6% of samples at Tabourie shared haplotypes with at least one other sample from this site. However, for the unstocked sites the respective values were very low (~1% Durras and ~2% for Corunna), which provides an estimate of natural background sharing. Given background sharing is around 1%, the values for the stocked sites could guide our estimates of the upper range of contribution rates for released prawns. Since eastern king prawn do not reproduce within estuaries, and rely upon a marine migratory phase to complete their life-cycle (Montgomery et al., 2007), isolation and co-ancestry can be excluded as possible causes for haplotype sharing. Our conclusion that shared haplotypes are due to kinship is further supported by the high level of haplotype sharing (more than 80%) occurring shortly after the first release. Pairwise comparison of PhiPT values for the temporal data indicated a significant difference for Wallagoot samples collected immediately after the release, which aligns with the high level of haplotype sharing shortly after the first release. Thus, the significant PhiPT values are assumed to arise from the inclusion of large family groups that leads to population differentiation (Wang, 2018). Finally, statistically fewer haplotypes at Wallagoot compared with the other estuaries reinforces the hypothesis of a high level of kinship at this site.

It is noteworthy that eastern king prawn collected from stocked estuaries following the second release event, when broodstock samples were available, haplotypes that were found within the hatchery broodstock were also detected in the stocked estuaries, although they were detected at low frequencies (1%–2%). The low detection rate may be due to the unknown number of missing broodstock samples. Moreover,

it has been reported there is a high variance in prawn fecundity between females in the hatchery (Knibb et al., 2014a), so it could be that the females contributing most to the stocked post-larvae were not present within the sample set. Importantly, by using inferred kinship groups, contributions of stocked animals to wild populations can still be determined despite the lack of genetic information from the broodstock (also seeTaylor et al., 2021).

There was a substantial difference in the proportion of shared haplotypes (and therefore assumed kinship and stocking signal), between Wallagoot and Tabourie. In particular, over the 5 months following the 2014 release, over 80% of eastern king prawn collected from Wallagoot appear to be stocked. This strong signal at Wallagoot can be attributed to the long-term closure of the estuary to the sea, creating a type of marine lake, limiting the emigration of the animals out to sea and eliminating natural recruitment and dilution of the proportion of restocked animals into the system. Given this, releasing prawns in similar closed estuaries may be more successful than releasing in more open estuaries.

Our results for Wallagoot highlight the potential impacts of stocking in severely recruitment limited estuaries, where stocked eastern king prawn can provide benefits to anglers for at least 5 months post-release, and these findings align with previous surveys of pilot release in these systems (Taylor, 2017). The other stocked site, Tabourie, maintained connectivity with the ocean leading up to both stocking events, probably allowing natural recruitment of wild eastern king prawn into the system. However, we still observed elevated levels of shared haplotypes (6%) at Tabourie compared to the reference estuaries (1%–2%). If we correct the proportion of eastern king prawn with shared haplotypes for background levels estimated in the reference estuaries, it appears the stocking could have contributed up to 5% of animals within this system.

Collectively, these findings show that eastern king prawn releases contribute to estuarine fisheries even where natural recruitment is occurring. Last, our results demonstrate that given the likely absence of broodstock information, the kinship approach is an effective method of identifying stocked animals.

Lastly, on the basis of our results, we reflect on the efficiency and cost effectiveness of molecular tagging vs. classical tagging for stock enhancement programs, especially those for crustacea. Historically, released fish/individuals were detected using different chemical markers, for example, otolith marking supported monitoring for black bream *Acanthopagrus butcheri* (Cottingham et al., 2020) and mulloway *Argyrosomus japonicus* (Taylor et al., 2021), or physical tags (elastomer tagging, coded wire tags) for trout species (*Salmo* and *Oncorhynchus* spp.) and red snapper *Lutjanus campechanus* (Hale and Gray, 1998; Brennan et al., 2007). For monitoring of released crustacea, these aforementioned historical methods are probably inappropriate because crustacea moult (losing any potentially stained or marked exoskeleton) and/or are too small during larval and post larval stages to accept wire tags/elastomer tagging. During the last decades, genetic tagging approaches have been used increasingly for monitoring the origin, family lines and kinship of the released animals (any species and any life stages) due to a range of advantages/considerations, not the least that individuals do not need to be tagged *per se* (they carry their own genetic tags) (Taylor et al., 2021). Bravington and Ward (2004) assessed the utility of DNA microsatellite genotyping and statistical approaches in stock enhancement programs for brown tiger prawns *P. esculentus*. The use of DNA microsatellites is still valid today, although newer cost equivalent methods such as SNPs are widely available (Premachandra et al., 2019). Should there be a high level of diversity for mitochondrial haplotypes (as we found for eastern king prawns), then use of mtDNA sequences can be a more cost effective but less technically demanding option than DNA microsatellites or SNPs, and indeed we found that mtDNA haplotypes well informed the present study as to the success of the prawn releases.

## 5 Conclusion

The present study reports several novel results and conclusions from an attempted prawn stock enhancement program (which in itself is a rather novel undertaking) and also novel results from population genetic assessments. Firstly, we conclude nearly all wild samples have unique mtDNA haplotypes, perhaps reflective of large population sizes. Second, notwithstanding the diversity, we conclude no significant differences among populations suggesting a panmictic stock along the NSW coast. Third, to compensate for incomplete broodstock records and samples, we used a novel approach of comparing the level of haplotype sharing to infer restocking success assuming shared haplotypes likely indicate siblings. The restocked estuaries had much higher frequencies of shared haplotypes than the unstocked estuaries suggestive of a stocking signal. Lastly, we conclude that the strongest stocking signal was at a location isolated from the sea, effectively a type of lake, and this novel knowledge may assist further stocking programs.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, ON804899–ON805769.

## Author contributions

HP performed the laboratory work, data analyses and participated in manuscript write up. WK contributed to the

data analysis, manuscript writing and project design. AB and MT provided samples and funding, project design, suggested analyses, revisions, editing of the manuscript.

## Funding

This project was funded by the NSW Recreational Fishing Saltwater Trust and samples were collected under permit P01/0059(A)-2.0 and ethics approval 14-01 by the NSW Department of Primary Industries.

## Acknowledgments

We would like to thank a number of people who involved in collecting and processing samples, Dylan Ven Der Meulen, Isabelle Thiebaud, Tristan New, Emma Simpson, James McLeod, Mitch Burns and Brooke McCartin. We appreciate discussions with Sankar Subramanian on initial drafts.

## Conflict of interest

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.975174/full#supplementary-material

## References

Becker, A., Lowry, M. B., and Taylor, M. D. (2018). Response of estuarine consumer communities following the stocking of a juvenile penaeid (*Penaeus plebejus*) over two consecutive years. *Fish. Manag. Ecol.* 25, 54–65. doi:10.1111/fme.12267

Blankenship, H. L., and Leber, K. M. (1995). A responsible approach to marine stock enhancement. *Am. Fish. Soc. Symp.* 15, 167–175.

Bravington, M. V., and Ward, R. D. (2004). Microsatellite DNA markers: Evaluating their potential for estimating the proportion of hatchery-reared offspring in a stock enhancement programme. *Mol. Ecol.* 13, 1287–1297. doi:10.1111/j.1365-294X.2004.02116.x

Brennan, N. P., Leber, K. M., and Blackburn, B. R. (2007). Use of coded-wire and visible implant elastomer tags for marine stock enhancement with juvenile red snapper *Lutjanus campechanus*. *Fish. Res.* 83, 90–97. doi:10.1016/j.fishres.2006.08.021

Chan, J. T. (2015). *Genetic analysis of the geographic structure of Australian eastern king prawns,* Penaeus (Melicertus) plebejus*, and implications for stock enhancement*. Sydney: Faculty of Science, University of NSW.

Chan, J. T., Sherwin, W. B., and Taylor, M. D. (2014). A tool for tracking genetic contributions of wild *Penaeus (Melicertus) plebejus* broodstock to hatchery populations. *Anim. Genet.* 45, 888–892. doi:10.1111/age.12212

Cothran, E. G., Juras, R., and Macijauskiene, V., 2005. Mitochondrial DNA D-loop sequence variation among 5 maternal lines of the Zemaitukai horse breed. *Genet. Mol. Biol.* 28, 677–681. doi:10.1590/S1415-47572005000500006

Cottingham, A., Hall, N. G., Loneragan, N. R., Jenkins, G. I., and Potter, I. C. (2020). Efficacy of restocking an estuarine-resident species demonstrated by long-term monitoring of cultured fish with alizarin complexone-stained otoliths. A case study. *Fish. Res.* 227, 105556. doi:10.1016/j.fishres.2020.105556

Davenport, J., Ekaratne, S. U. K., Walgama, S. A., Lee, D., and Hills, J. M. (1999). Successful stock enhancement of a lagoon prawn fishery at Rekawa, Sri Lanka using cultured post-larvae of penaeid shrimp. *Aquaculture* 180, 65–78. doi:10.1016/S0044-8486(99)00141-6

Diniz, F. M., Maclean, N., Ogawa, M., Cintra, I. H. A., and Bentzen, P. (2005). The hypervariable domain of the mitochondrial control region in Atlantic spiny lobsters and its potential as a marker for investigating phylogeographic structuring. *Mar. Biotechnol.* 7, 462–473. doi:10.1007/s10126-004-4062-5

Everett, J. D., van Sebille, E., Taylor, M. D., Suthers, I. M., Setio, C., Cetina-Heredia, P., et al. (2017). Dispersal of Eastern King Prawn larvae in a Western boundary current: New insights from particle tracking. *Fish. Oceanogr.* 26, 513–525. doi:10.1111/fog.12213

Hale, R. S., and Gray, J. H. (1998). Retention and detection of coded wire tags and elastomer tags in trout. *N. Am. J. Fish. Manag.* 18, 197–201. doi:10.1577/1548-8675(1998)018<0197:radocw>2.0.co;2

Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Hamasaki, K., and Kitada, S. (2006). A review of kuruma prawn *Penaeus japonicus* stock enhancement in Japan. *Fish. Res.* 80, 80–90. doi:10.1016/j.fishres.2006.03.018

Helidoniotis, F., O'Neill, M. F., and Taylor, M. (2020). *Stock assessment of eastern king prawn (*Melicertus plebejus*)*. Queensland, State of Queensland, Australia: Department of Agriculture and Fisheries.

Hsu, T.-H., Huang, C.-W., Lin, C.-H., Lee, H.-T., and Pan, C.-Y. (2020). Tracing the origin of fish without hatchery information: Genetic management of stock enhancement for mangrove red snapper (Lutjanus argentimaculatus) in taiwan. *Fish. Aquat. Sci.* 23, 13. doi:10.1186/s41240-020-00156-9

Kelemec, J. A., and Smith, I. R. (1980). Induced ovarian development and spawning of penaeus-plebejus in a recirculating laboratory tank after unilateral eyestalk enucleation. *Aquaculture* 21, 55–62. doi:10.1016/0044-8486(80)90125-8

Kitada, S. (2018). Economic, ecological and genetic impacts of marine stock enhancement and sea ranching: A systematic review. *Fish. Fish.* 19, 511–532. doi:10.1111/faf.12271

Kitada, S. (2020). Lessons from Japan marine stock enhancement and sea ranching programmes over 100 years. *Rev. Aquac.* 12, raq.12418–1969. doi:10.1111/raq.12418

Knibb, W., Giang, C. T., Premachandra, H. K. A., Ninh, N. H., and Domínguez, B. C. (2020). Feasible options to restore genetic variation in hatchery stocks of the globally important farmed shrimp species, Litopenaeus vannamei. *Aquaculture* 518, 734823. doi:10.1016/j.aquaculture.2019.734823

Knibb, W., Quinn, J., Lamont, R., Whatmore, P., Nguyen, N. H., and Remilton, C. (2014a). Reproductive behaviour of captive *Fenneropenaeus merguiensis*: Evidence for monogamy and high between family variances for offspring number. *Aquaculture* 426, 60–65. doi:10.1016/j.aquaculture.2014.01.025

Knibb, W., Whatmore, P., Lamont, R., Quinn, J., Powell, D., Elizur, A., et al. (2014b). Can genetic diversity be maintained in long term mass selected populations without pedigree information? — a case study using banana shrimp *Fenneropenaeus merguiensis*. *Aquaculture* 428-429, 71–78. doi:10.1016/j.aquaculture.2014.02.026

Librado, P., and Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452. doi:10.1093/bioinformatics/btp187

Liu, J. Y. (1990). Resource enhancement of Chinese shrimp, *Penaeus orientalis*. *Bull. Mar. Sci.* 47, 124–133.

Liu, Q., Cui, F., Hu, P. F., Yi, G. T., Ge, Y. W., Liu, W. L., et al. (2018). Using of microsatellite DNA profiling to identify hatchery-reared seed and assess potential genetic risks associated with large-scale release of swimming crab *Portunus trituberculatus* in Panjin, China. *Fish. Res.* 207, 187–196. doi:10.1016/j.fishres.2018.05.003

Loneragan, N. R., Ye, Y. M., Kenyon, R. A., and Haywood, M. D. E. (2006). New directions for research in prawn (shrimp) stock enhancement and the use of models in providing directions for research. *Fish. Res.* 80, 91–100. doi:10.1016/j.fishres.2006.03.014

Lorenzen, K., Leber, K. M., and Blankenship, H. L. (2010). Responsible approach to marine stock enhancement: An update. *Rev. Fish. Sci.* 18, 189–210. doi:10.1080/10641262.2010.491564

McMillen-Jackson, A. L., and Bert, T. M. (2003). Disparate patterns of population genetic structure and population history in two sympatric penaeid shrimp species (*Farfantepenaeus aztecus* and *Litopenaeus setiferus*) in the eastern United States. *Mol. Ecol.* 12, 2895–2905. doi:10.1046/j.1365-294X.2003.01955.x

McMillen-Jackson, A. L., and Bert, T. M. (2004). Genetic diversity in the mtDNA control region and population structure in the pink shrimp *Farfantepenaeus duorarum*. *J. Crustac. Biol.* 24, 101–109. doi:10.1651/C-2372

Molony, B. W., Lenanton, R., Jackson, G., and Norriss, J. (2003). Stock enhancement as a fisheries management tool. *Rev. Fish. Biol. Fish.* 13, 409–432. doi:10.1007/s11160-004-1886-z

Montgomery, S. S., Courtney, A. J., Blount, C., Stewart, J., Die, D. J., Cosgrove, M., et al. (2007). Patterns in the distribution and abundance of female eastern king prawns, *Melicertus plebejus* (Hess, 1865), capable of spawning and reproductive potential in waters off eastern Australia. *Fish. Res.* 88, 80–87. doi:10.1016/j.fishres.2007.08.002

Montgomery, S. S. (1990). Movements of juvenile eastern king prawns, *penaeus-plebejus*, and identification of stock along the east-coast of Australia. *Fish. Res.* 9, 189–208. doi:10.1016/S0165-7836(05)80001-3

Niwa, Y., Nakazawa, A., Margulies, D., Scholey, V. P., Wexler, J. B., and Chow, S. (2003). Genetic monitoring for spawning ecology of captive yellowfin tuna (*Thunnus albacares*) using mitochondrial DNA variation. *Aquaculture* 218, 387–395. doi:10.1016/S0044-8486(03)00015-2

Obata, Y., Imai, H., Kitakado, T., Hamasaki, K., and Kitada, S. (2006). The contribution of stocked mud crabs Scylla paramamosain to commercial catches in Japan, estimated using a genetic stock identification technique. *Fish. Res.* 80, 113–121. doi:10.1016/j.fishres.2006.03.016

Peakall, R., and Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28, 2537–2539. doi:10.1093/bioinformatics/bts460

Premachandra, H. K. A., La Cruz, F. L., Takeuchi, Y., Miller, A., Fielder, S., O'Connor, W., et al. (2017). Genomic DNA variation confirmed *Seriola lalandi* comprises three different populations in the Pacific, but with recent divergence. *Sci. Rep.* 7, 9386. doi:10.1038/s41598-017-07419-x

Premachandra, H. K. A., Nguyen, N. H., and Knibb, W. (2019). Effectiveness of SNPs for parentage and sibship assessment in polygamous yellowtail kingfish *Seriola lalandi*. *Aquaculture* 499, 24–31. doi:10.1016/j.aquaculture.2018.09.022

Reid, D. D., and Montgomery, S. S. (2005). Creel survey based estimation of recreational harvest of penaeid prawns in four southeastern Australian estuaries and comparison with commercial catches. *Fish. Res.* 74, 169–185. doi:10.1016/j.fishres.2005.03.007

Roy, P. S., Williams, R. J., Jones, A. R., Yassini, I., Gibbs, P. J., Coates, B., et al. (2001). Structure and function of south-east Australian estuaries. *Estuar. Coast. Shelf Sci.* 53, 351–384. doi:10.1006/ecss.2001.0796

Ruello, N. V. (1975). Geographical distribution, growth and breeding migration of the eastern Australian king prawn *Penaeus plebejus* hess. *Mar. Freshw. Res.* 26, 343–354. doi:10.1071/mf9750343

Serajuddin, M., Bano, F., Awasthi, M., Gupta, P., and Kumar, G. (2018). Marine stock enhancement in India: Current status and future prospects, in *Marine ecology - biotic and abiotic interactions*: M. Türkoğlu, U. Önal, and A. Ismen (London: IntechOpen), 175–195. doi:10.5772/intechopen.75175

Setio, C. (2016). *Evaluating the stock enhancement of eastern king prawns into a south-east Australian estuary, Evolution and Ecology Research Centre*. Sydney, Australia: University of New South Wales.

Stoneking, M. (2000). Hypervariable sites in the mtDNA control region are mutational hotspots. *Am. J. Hum. Genet.* 67, 1029–1032. doi:10.1086/303092

Taiarui, M., Foale, S., Bambridge, T., and Sheaves, M. (2019). Is stock enhancement the best option to manage fisheries? A case study from taiarapu (French polynesia). *Mar. Policy* 104, 1–11. doi:10.1016/j.marpol.2019.02.026

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11 molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi:10.1093/molbev/msab120

Taylor, M. D., and Johnson, D. D. (2021). Connectivity between a spatial management network and a multi-jurisdictional ocean trawl fishery. *Ocean. Coast. Manag.* 210, 105691. doi:10.1016/j.ocecoaman.2021.105691

Taylor, M. D., Palmer, P. J., Fielder, D. S., and Suthers, I. M. (2005). Responsible estuarine finfish stock enhancement: An Australian perspective. *J. Fish. Biol.* 67, 299–331. doi:10.1111/j.0022-1112.2005.00809.x

Taylor, M. D. (2017). Preliminary evaluation of the costs and benefits of prawn stocking to enhance recreational fisheries in recruitment limited estuaries. *Fish. Res.* 186, 478–487. doi:10.1016/j.fishres.2016.05.030

Taylor, M. D., Premachandra, H. K. A., Hurwood, D. A., Dammannagoda, S. T., Chan, K. H., Mather, P. B., et al. (2021). Genetic evaluation of the unknown contribution of stocked fish in angler catches: A case study using mulloway *Argyrosomus japonicus*. *Bull. Mar. Sci.* 97, 599–614. doi:10.5343/bms.2020.0050

Teixeira, H., Rodriguez-Echeverria, S., and Nabais, C. (2014). Genetic diversity and differentiation of *Juniperus thurifera* in Spain and Morocco as determined by SSR. *Plos One* 9, e88996. doi:10.1371/journal.pone.0088996

Wang, J. L. (2018). Effects of sampling close relatives on some elementary population genetics analyses. *Mol. Ecol. Resour.* 18, 41–54. doi:10.1111/1755-0998.12708

Wang, M. S., Wang, W. J., Xiao, G. X., Liu, K. F., Hu, Y. L., Tian, T., et al. (2016). Genetic diversity analysis of spawner and recaptured populations of Chinese shrimp (*Fenneropenaeus chinensis*) during stock enhancement in the Bohai Bay based on an SSR marker. *Acta Oceanol. Sin.* 35, 51–56. doi:10.1007/s13131-016-0830-0

Wang, Q. Y., Zhuang, Z. M., Deng, J. Y., and Ye, Y. M. (2006). Stock enhancement and translocation of the shrimp *Penaeus chinensis* in China. *Fish. Res.* 80, 67–79. doi:10.1016/j.fishres.2006.03.015

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1081246



OPEN ACCESS

EDITED BY
Mehar S. Khatkar,
The University of Sydney, Australia

REVIEWED BY
Shengjie Ren,
Queensland University of Technology, Australia
Tao Zhou,
Xiamen University, China

*CORRESPONDENCE
Nguyen Hong Nguyen,
nnguyen@usc.edu.au
Nguyen Van Sang,
sangnv.ria2@mard.gov.vn

†These authors have contributed equally to this work

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

RECEIVED 27 October 2022
ACCEPTED 06 December 2022
PUBLISHED 04 January 2023

CITATION
Vu NT, Phuc TH, Nguyen NH and Van Sang N (2023), Effects of common full-sib families on accuracy of genomic prediction for tagging weight in striped catfish *Pangasianodon hypophthalmus*. *Front. Genet.* 13:1081246. doi: 10.3389/fgene.2022.1081246



# Effects of common full-sib families on accuracy of genomic prediction for tagging weight in striped catfish *Pangasianodon hypophthalmus*

Nguyen Thanh Vu[1,2,3], Tran Huu Phuc[3†], Nguyen Hong Nguyen[1,2*] and Nguyen Van Sang[3*†]

[1]School of Science, Technology and Engineering, University of the Sunshine Coast, Sippy Downs, QLD, Australia, [2]Center for Bio-Innovation, University of the Sunshine Coast, Maroochydore, QLD, Australia, [3]Research Institute for Aquaculture No. 2, Ho Chi Minh City, Vietnam

Common full-sib families ($c^2$) make up a substantial proportion of total phenotypic variation in traits of commercial importance in aquaculture species and omission or inclusion of the $c^2$ resulted in possible changes in genetic parameter estimates and re-ranking of estimated breeding values. However, the impacts of common full-sib families on accuracy of genomic prediction for commercial traits of economic importance are not well known in many species, including aquatic animals. This research explored the impacts of common full-sib families on accuracy of genomic prediction for tagging weight in a population of striped catfish comprising 11,918 fish traced back to the base population (four generations), in which 560 individuals had genotype records of 14,154 SNPs. Our single step genomic best linear unbiased prediction (ssGLBUP) showed that the accuracy of genomic prediction for tagging weight was reduced by 96.5%–130.3% when the common full-sib families were included in statistical models. The reduction in the prediction accuracy was to a smaller extent in multivariate analysis than in univariate models. Imputation of missing genotypes somewhat reduced the upward biases in the prediction accuracy for tagging weight. It is therefore suggested that genomic evaluation models for traits recorded during the early phase of growth development should account for the common full-sib families to minimise possible biases in the accuracy of genomic prediction and hence, selection response.

KEYWORDS

**genetic improvement, genomic selection, growth traits, non-additive genetic estimates and accuracy of selection response, genetic lines**

# 1 Introduction

In aquaculture species, common full-sib families ($c^2$) are a result of separate family rearing of about one to 3 months until larvae reach a suitable size for physical tagging (e.g., 10–20 g in fish or 2–5 g in shrimps). The $c^2$, also known as non-additive genetic components, include both common environmental and maternal effects or possibly dominance, accounting for a significant proportion of total phenotypic variations, ranging from 5% to 55% for growth-related traits in fish (Hamzah et al., 2017; Vu et al., 2019b; Bosworth et al., 2020), crustacean (Nguyen et al., 2020a; Sang N. V. et al., 2020), and mollusc (Sang V. V. et al., 2020). A meta-analysis of 45 studies available in the literature across aquaculture species showed that the mean $c^2$ value is about 10% for harvest body weight (Nguyen, 2021). Omission of the $c^2$ resulted in overestimation of heritability by 9%–45% in red tilapia *Oreochromis* spp. (Nguyen et al., 2017; Sukhavachana et al., 2019) or giant freshwater prawn *Macrobrachium rosenbergii* (Luan et al., 2012; Phuc et al., 2021). The estimates of common full-sib families were substantially larger for traits recorded during the early stage of growth development than those measured at harvest. For example, the $c^2$ values were estimated at .37 for tagging weight vs. .21 for harvest body weight in striped catfish *Pangasianodon hypophthalmus* (Vu et al., 2019b). However, to date, the impacts of common full-sib families on genomic prediction accuracy have not been reported in any aquaculture species, including striped catfish *P. hypophthalmus*.

Current genomic evaluation models used to analyse traits of commercial importance in aquaculture species include only genomic and phenotypic data or combined with pedigree information (e.g., single-step GBLUP). Under these models, the prediction accuracies for body traits (e.g., weight, length) at harvest were moderate to high, ranging from .38 to .89 (Houston et al., 2020). The prediction accuracies for early growth were .33 in common carp *Cyprinus carpio* (Palaiokostas et al., 2018) and .67 for Pacific oysters *Magellana gigas* (Gutierrez et al., 2018). The prediction accuracies for meat quality traits fall within a range of .59–.62 for raw and cooked colour of banana shrimp *Fenneropenaeus merguiensis* (Nguyen et al., 2020b) and .19–.20 for fillet yield and firmness in rainbow trout *Oncorhynchus mykiss* (Al-Tobasei et al., 2021). To date, there is no or limited published information regarding the utilisation of genomic data to assess predictive performance of any statistical methods for tagging weight (i.e., early growth trait) in important aquaculture species.

Almost all studies in aquaculture have employed genomic best linear unbiased prediction (GBLUP), single step- GBLUP (ssGBLUP) or Bayesian methods (Allal and Nguyen, 2022). The Bayesian methods provide flexibility to model different variance distributions of SNPs and can outperform BLUP method (e.g., GBLUP or ssGBLUP) especially for traits under control by genes with large and moderate effects (van den Berg et al., 2015). However, computation of Bayesian methods is highly demanding, and hence, BLUP-family methods have been widely used in practical breeding programs, especially for traits whose variation is of polygenic nature due to many genes, each with very small effects. Recent studies have employed machine and deep learning and obtained higher accuracies for a range of traits than linear (GBLUP) and non-linear Bayesian methods (Yin et al., 2020; Montesinos-López et al., 2021). Regardless of statistical methods used, imputation of missing genotypes or imputation from a low to high density SNP arrays or from commercial SNP arrays to whole genome sequence improved the prediction accuracy for complex traits (Kjetså et al., 2020). Multivariate analysis also slightly increased the prediction accuracy for grain yield in wheat *Triticum aestivum L.* (Sandhu et al., 2021) or cassava *Manihot esculenta* Crantz (Okeke et al., 2017), although its benefits depend on statistical models used (Montesinos-López et al., 2020) or characteristics of datasets and specifically genetic architecture of traits (Gianola and Fernando, 2020). Recent studies have also reported advantages of including functional variants identified from genome-wide associations analysis (GWAS) in prediction models to improve the accuracy of genomic estimated breeding values for growth traits under chronic thermal stress in rainbow trout *O. mykiss* (Yoshida and Yáñez, 2021). In this regard, published information is not available for tagging weight, especially in striped catfish—an important aquaculture species that contributes significantly to the national economies of many countries in Asia, such as Bangladesh, Malaysia, Thailand, Vietnam.

Therefore, this study was set out to test three major hypotheses: 1) omission of the common full-sib families ($c^2$) from statistical models can result in upward biases in genomic prediction accuracy for tagging weight, 2) imputation of missing genotypes can improve the predictive performance of ssGBLUP, and 3) multi-trait genomic evaluation can lessen the overestimation of the prediction accuracy when the common full-sib families were omitted. Ultimately, the study attempted to explore possibilities for the application of genomic selection for early growth traits in striped catfish.

# 2 Materials and methods

## 2.1 Source of genetic materials

This study included 11,918 fish, which are offspring of 434 females and 278 males in a full pedigree traced back to the base population. The experimental fish were produced between 2015 and 2020, following a semi-nested mating design with a ratio of one male to one or two females (Van Sang et al., 2012). Induced breeding was practised using HCG (Human Chorionic Gonadotropin) hormone with 4 doses (300, 600, 1,200, and

**TABLE 1 Descriptive statistics for tagging weight of striped catfish data over 2 generations.**

| Index | G0-resistance | G1-resistance | G1-growth | All generations |
|---|---|---|---|---|
| Observation | 4937 | 5224 | 1757 | 11918 |
| Number of fish per family | 27.6 (10–86) | 31.3 (13–87) | 18.9 (11–72) | 27.1 (10–87) |
| Weight (g) | 23.9 ± 11.7 | 20.8 ± 11.7 | 25.0 ± 15.2 | 22.7 ± 12.4 |
| CV of weight (%) | 48.8 | 55.9 | 60.7 | 54.4 |
| Age in day (min—max) | 195.0 (149–208) | 148.6 (132–180) | 178.6 (132–222) | 172.2 (132–222) |
| No. batches | 9 | 4 | 7 | 19 |
| No. of sire | 107 | 99 | 72 | 272 |
| No. of dam | 177 | 167 | 90 | 428 |
| No. of families | 179 | 167 | 93 | 439 |

3500 UI). Also note that there are different induction practices regarding doses and types of hormones used, e.g., HCG (Bui et al., 2010) or Suprefact (Samorn, 2007). After striping and ferilizing, eggs were incubated in net jars mounted in a 5 $m^3$ composite tank. After hatching, fry of each family was reared in a separate fibreglass tank (1.5 $m^3$) for about 3 weeks. Then a random sample of about 500 fry per family were transferred to a net hapa installed in earthen ponds to raise to fingerling size of about 20 g for physical tagging, using Passive Integrated Transponder (PIT). One family was kept in a single hapa net. Three feeding strategies were applied for different rearing periods: no feeding before hatching (0–24 h), Artemia (day 1 to day 3) and Moina (day 4 to day 7) together with fish flake (day 8 to day 15, the foods were made of small size before feeding) were used for tank rearing period before 15 days (Vu et al., 2019b) at a maximum fish uptake and only pellet feed was used during 2 months rearing in earthen pond at a rate of 5% fish biomass. The water was exchanged 50% daily when fish were kept in tank and once per week in pond. In each generation, approximately 200 fish were randomly sampled from each family for PIT tagging. And a half of each family was used for growth testing in the mainstream selection program for increased harvest body weight (Vu et al., 2019b) and another half was sent to concrete tanks for pathogen challenge test to select for increased resistance to *Edwardsiella ictaluri*, a bacterial disease that has caused severe mortality loss during larval and fingerling rearing stages in striped catfish (Vu et al., 2019a). Due to the high mortalities observed after tagging and conditioning, there was a smaller number of fish per family retained for the main challenge test, around 27 fish/family (Table 1).

## 2.2 Trait(s) studied

At tagging, weight of individual fish was recorded using a digital scale with a precision to .1 g. In 2015, 4,937 fish had tagging weight and in the latest generation in 2020, the number of fish with tag weight involved in the pathogen challenge experiment and growth performance testing were 5,224 and 1,757, respectively. In total, there were 11,918 individual data records used in our statistical analysis to assess genomic prediction accuracies. However, due to our limited funding, only a random sample of 560 fish from 40 families in the latest generation (2020) was sequenced to obtain genotype data for our analysis in this study.

## 2.3 Genotype

DNA samples of 560 fish (offspring of 40 females and 31 males) were sent to a commercial service provider in Canberra, Australia for genotyping by sequencing, using Diversity Arrays Technology (DArTseq™). DArTseq™ represents a combination of genome complexity reduction methods and high throughput sequencing platforms (Kilian et al., 2012). A detailed description regarding selections of restricted enzymes, PCR reactions, library preparations and sequencing is given in our earlier studies (Nguyen et al., 2018a; Nguyen et al., 2018b; Nguyen et al., 2020b; Vu et al., 2020). Briefly, sequences generated from each lane were processed using proprietary DArTseq pipelines. Approximately 2,000,000 sequences per barcode/sample were identified and used for variant (SNP) calling. SNP calling was conducted in the DArTsoft14, using DART PL's C++ algorithm. Calling quality was assured by high average read depth (averaging 60 reads per locus). Furthermore, when multiple polymorphisms were detected on DNA fragments (mostly 75 bp), a single SNP was randomly chosen to avoid linkage disequilibrium between loci. After quality control (QC), we obtained 14,154 SNPs across 560 samples.

## 2.4 Statistical analysis

The missing genotypes (about 10.0%) were imputed using AlphaFamImpute (involving 560 individuals fish and 14,154 SNPs) which takes into account of the pedigree

relationships (Whalen et al., 2020). Single-step genomic best linear unbiased prediction (ssGBLUP) method was used to assess genomic prediction accuracy for tag weight. The linear mixed model is written in a matrix notation as follows:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{Wc} + \boldsymbol{e} \quad (1)$$

where

$\boldsymbol{y}$ is the observations related to individual records of each fish.

$\boldsymbol{X}$ is the design matrix related to fixed estimates ($\boldsymbol{b}$) that included generation (1–3) and spawning batch. Age from birth to tagging was also fitted as a linear covariate.

$\boldsymbol{Z}$ and $\boldsymbol{W}$ are the design matrices related to the additive genetic effects $\boldsymbol{u} \sim \boldsymbol{(0, H\sigma^2 g)}$ and common full-sib groups $\boldsymbol{c} \sim \boldsymbol{(0, I\sigma^2 c)}$. The random terms fitted in the model were the additive genetics of individual fish and the common full-sib families. LogLikelihood Ratio Test (LRT) showed that the common full-sib effects were statistically significant for tag weight (Chi-square value with one degree of freedom ranged from 2.3 to 6.2, $p < .05$ to .001). $\boldsymbol{e} \sim \boldsymbol{(0, I\sigma^2 e)}$ is the error term in the model. Where $\boldsymbol{I}$ is the indentity matrix, $\boldsymbol{H}$ is described as below. $\boldsymbol{\sigma^2 g}$, $\boldsymbol{\sigma^2 c}$, $\boldsymbol{\sigma^2 e}$ are corresponding genetic, common environmental and residual variances.

Our ssGBLUP analysis was conducted in AIREMLf90 of the BLUPF90 package (Misztal et al., 2018). The ssGBLUP is an advanced version of GBLUP that blended numerator relationship ($\boldsymbol{A}$) and kinship ($\boldsymbol{G}$) matrices into a realised $\boldsymbol{H}$ matrix (Eq. 2), where $\boldsymbol{A}$ was calculated from the pedigree and $\boldsymbol{G}$ was computed from 14,154 SNPs. ssGBLUP uses the blended matrix combining both pedigree information and genotype data and hence, is deemed more powerful than GBLUP.

$$\boldsymbol{H^{-1}} = \boldsymbol{A^{-1}} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G^{-1} - A_{22}^{-1}} \end{bmatrix} \quad (2)$$

The model for single step GWAS expressed as below [also see Aguilar et al. (2019)]:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Z_i a_i} + \boldsymbol{u} + \boldsymbol{e} \quad (3)$$

where $Z_i$ is a vector of SNP values (i.e., 0, 1 or 2), $a_i$ is the effect of the $i$th SNP, $u$ is the vector of breeding values obtained from single step analysis from Eq. 1. Here,

$$\mathbf{var}(\boldsymbol{u}) = \frac{\boldsymbol{ZZ'}}{\sum(\boldsymbol{p_i}(\boldsymbol{1} - \boldsymbol{p_i}))^2} = \boldsymbol{G\sigma_u^2} \quad (4)$$

with $p_i$ is the frequency of $i$th SNP, $\sigma_u^2$ and $\sigma_e^2$ are assumed known and $y$, $X$, $b$, $Z$, $G$, and $e$ are described as above. Analysis of ssGWAS was accomplished by three sub-programs, including blupf90 (computation of genomic breeding values), pregsf90 (derivation of the H matrix) and postgsf90 (estimation of the SNP effect, $p$-values and plotting). The pre-selected SNP panels after GWAS analysis were based on a significant probability of less than .00001 for each of the 25 running sets. Finally, analysis of ssGBLUP genomic prediction were performed using only the highly significant SNPs. The model that omitted the common full-sib effects ($c^2$) was the same as Model 1, except that the "Wc" term or "full-sibs" effect was not included.

The predictive performance (or prediction accuracy) of ssGBLUP was evaluated using 5-fold cross validation over five replications. This involved the random division of the phenotypic data into 5 subsets (each with 2383–2384 observations). Then the breeding value of one set was predicted based on the data from the other four subsets. In the five fold cross-validation, the process was repeated 5 times and thus, there were 25 runs in total. The genomic prediction accuracy was defined as the correlations between the predicted breeding values and actual phenotypes ($r_{y,\hat{y}}$) divided by the square root of the trait heritability. The trait heritability was estimated using the AIREML algorithm in the AIREMLF90 of the BLUPF90 family package. The method assumed normal distribution of the variance components for the traits studied; they were the observed heritability for the trait studied. The correlations ($r_{y,\hat{y}}$) were determined as the average value after five-fold cross-validation with 5 repetitions. All single trait analyses were performed in AIREMLf90. Regarding the bivariate analysis, tag weight was co-analysed with survival time (i.e., days that the animals were still alive after the challenge test experiment). The bivariate model was also performed in AIREMLf90. In addition, we analysed the two-trait model [tag weight and survival time (Vu et al., 2021)], using Gibb Sampling method in THRGIBBF1f90 (Tsuruta and Misztal, 2006). In both packages, the bivariate model was the same as Eq. 3 above. In the Gibb sampling, we used 200,000/20,000 and 1,000,000/200,000 as total-cycle/burn-in steps for the univariate and bivariate analyses, respectively. After each Gibbs sampling analysis, results obtained from all the samples were visualised using time series plots of postgibbsf90 program (https://masuday.github.io/blupf90_tutorial/vc_gs.html) to define the stability of variances, and only samples displaying stabilised variances were used to calculate heritability and/or genetic parameters. The prediction accuracies obtained from AIREMLf90 were almost identical to those obtained from THRGIBBF1f90. Thus, only the estimates from the latter analysis were presented in this study. Finally, we performed pedigree-based PBLUP analysis and single-step genome-wide association study (GWAS) in combination of ssGBLUP to better understand the predictive capacity of our statistical models used to analyse tag weight. These analyses used AIREMLf90 and THRGIBBF1f90 packages (Misztal et al., 2018).

Finally, we calculated correlation of EBV for tagging weight between the two statistical models (with and without the common full-sib families) to examine re-ranking effects, i.e., re-ranking of breeding candidates based on their EBVs for tagging weight in the selection program for striped catfish.

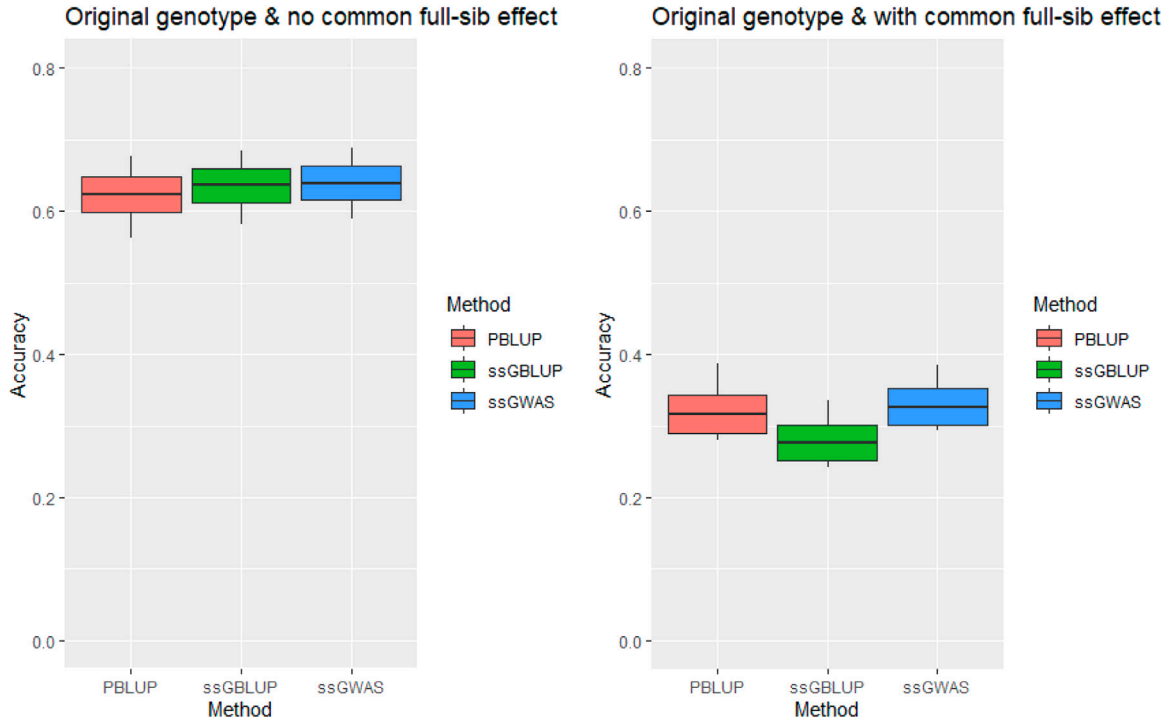



FIGURE 1
Accuracy of genomic prediction for tagging weight without/with common full-sib effect ($c^2$) using original genotype under AI-REML algorithm. Middle line of the box is mean accuracy; top and bottom lines of the box is accuracy ± one standard deviation. End points of vertical line represent minimum and maximum values.

# 3 Results

## 3.1 Trait characteristics

The average tag weight of the population was 22.7 ± 12.4 g (Table 1). The tag weight in the first generation (G1, produced in 2019) was slightly lower than that of the base population (G0, produced in 2015) as the animals were tagged at an earlier age (149 vs. 195 d). Despite our efforts to produce all families within a short period in G1 (4–7 spawning batches), the coefficient of variation in the tag weight was somewhat greater in this generation than in the base population (55.9%–60.7% vs. 48.8%). Note that only the animals of generation 1 (560 individuals) had genome sequence and genotype (SNPs) data. The average tag weight of these animals was 23.2 ± 13.0 g.

The heritability ($h^2$) for tag weight was high (.72–.74) when the common full-sib estimate ($c^2$) was omitted from our models: PBLUP, ssGBLUP and ssGWAS (Supplementary Table S1). The $h^2$ estimate obtained from the full models that also included the $c^2$ estimate was reduced to .15, .08, and .14 for PBLUP, ssGBLUP, and ssGWAS, respectively. The corresponding $c^2$ estimates were .71, .74, and .72 (Supplementary Table S1).

## 3.2 Accuracy of genomic prediction with and without common full-sib effect ($c^2$)

The genomic prediction accuracy for tag weight was high (.636) when the $c^2$ estimates were omitted from our statistical model. However, the accuracy was significantly reduced to .276 in the ssGBLUP model that also included the $c^2$ estimates (Figure 1). In other words, omission of the $c^2$ resulted in loss of the prediction accuracy by .278–.334 (or 80.7–105.3%).

## 3.3 Original vs. imputed data using the full model

Imputation of missing genotypes alleviated the upward bias in the prediction accuracy for tag weight when the $c^2$ estimates were fitted in statistical models of our analysis (Figure 2). The accuracy obtained from the full ssGBLUP model that included the $c^2$ estimates was .311 when the imputed genotype was analysed as compared with .276 of the original data. This means that imputation improved the prediction accuracy for tag weight by .035 (or 12.8%) (also see Supplementary Table S2).

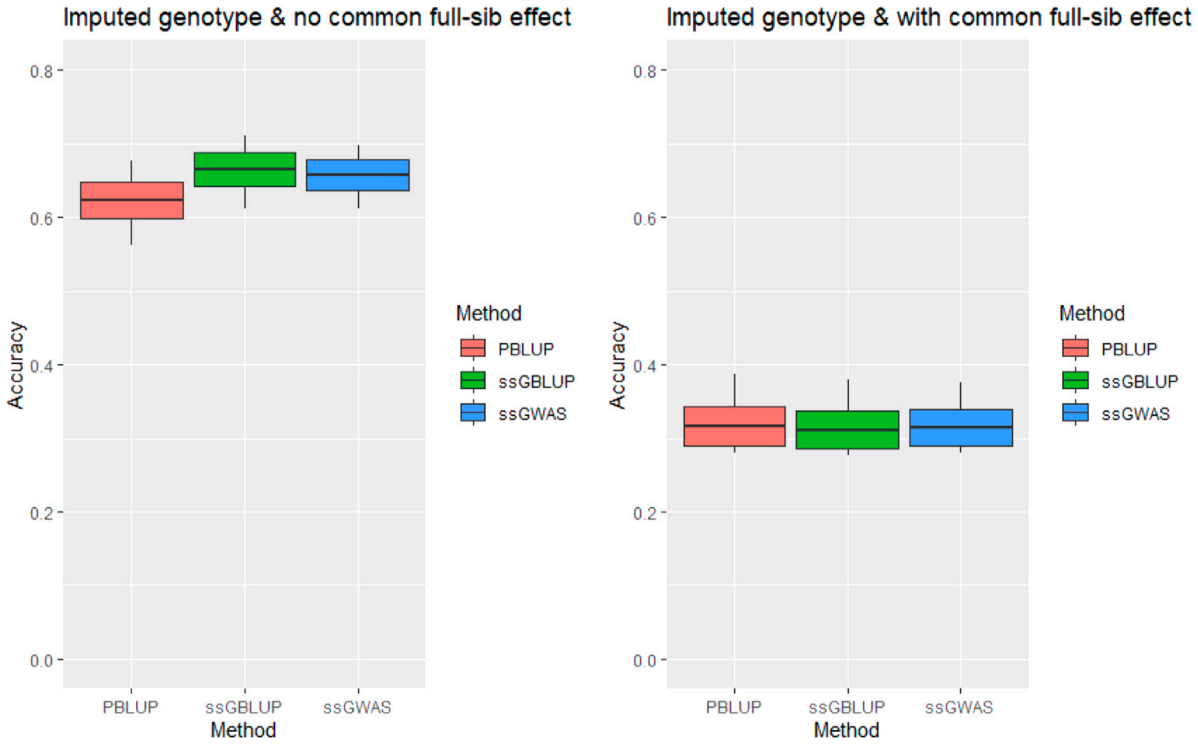



FIGURE 2
Accuracy of genomic prediction for tagging weight without/with common full-sib effect ($c^2$) using imputed genotype under AI-REML algorithm. Middle line of the box is mean accuracy; top and bottom lines of the box is accuracy ± one standard deviation. End points of vertical line represent minimum and maximum values.

**TABLE 2 Genomic prediction accuracy from multivariate models in AIREMLf90 and THRGIBBS1f90, using original (un-imputed) genotypes.**

| Method | AIREMLf90 | | | THRGIBBS1f90 | | |
|---|---|---|---|---|---|---|
| | Without $c^2$ | With $c^2$ | Difference (%) | Without $c^2$ | With $c^2$ | Difference (%) |
| PBLUP | .6272 ± .025 | .3471 ± .025 | 80.7 | .6227 ± .025 | .3446 ± .025 | 80.7 |
| ssGBLUP | .6371 ± .023 | .3067 ± .023 | 107.7 | .6360 ± .023 | .3098 ± .022 | 105.3 |
| ssGWAS | .6391 ± .023 | .3264 ± .026 | 95.8 | .6392 ± .023 | .3451 ± .026 | 85.2 |

## 3.4 Multi-trait analysis using the full model

Bivariate analysis involving tag weight and a disease resistance trait (i.e., survival time) improved the prediction accuracy by .031 (or 11.2%) relative to the univariate ssGBLUP model (.307 vs. .276). The two-trait analysis also reduced biases in the prediction accuracy for tag weight when the $c^2$ were included in our models ($r$ = .3098 for the two-trait model with the $c^2$ estimates vs. .630 without the $c^2$). Regardless of the inclusion or exclusion of the $c^2$, the prediction accuracies were similar between AI-REML and Gibb sampling methods either when original genotypes (Table 2) or imputed genotypes (Table 3) were analysed.

## 3.5 ssGWAS in combination with ssGBLUP

The inclusion of highly significant markers (471 SNPs) slightly increased the prediction accuracy for tag weight relative to ssGBLUP (Table 2). However, it had little impacts on the upward biases in the prediction accuracy when the common full-sib families were omitted from our univariate

**TABLE 3 Genomic prediction accuracy from multivariate models in AIREMLf90 and THRGIBBS1f90, using imputed genotypes.**

| Method | AIREMLf90 | | | THRGIBBS1f90 | | |
|---|---|---|---|---|---|---|
| | Without $c^2$ | With $c^2$ | Difference (%) | Without $c^2$ | With $c^2$ | Difference (%) |
| PBLUP | .6272 ± .025 | .3471 ± .025 | 80.7 | .6227 ± .025 | .3446 ± .025 | 80.7 |
| ssGBLUP | .6644 ± .023 | .3323 ± .026 | 99.9 | .6658 ± .023 | .3317 ± .024 | 100.7 |
| ssGWAS | .6571 ± .022 | .3352 ± .023 | 96.0 | .6578 ± .021 | .3331 ± .021 | 97.5 |

(Figures 1, 2) and multi-variate analyses (Table 3), using either linear mixed model or threshold Gibb sampling methods.

## 3.6 Re-ranking effects

To examine the impact of the common full-sib families on re-ranking effects, we calculated correlation of EBV for tagging weight between the two models (with the presence and absence of the common full-sibs effect). The Pearson correlation coefficient ranged from .30 to .62 (Supplementary Table S3), suggesting potential re-ranking effects of selection candidates when the $c^2$ effects were not included in genomic evaluation models for tagging weight of striped catfish.

# 4 Discussion

In the present study we attempted to address five major questions which are worth considering before initiating genomic selection program for early growth in striped catfish as well as other aquaculture species of economic importance.

## 4.1 Should genomic selection be practised for tagging weight?

The prediction accuracy for tagging weight was high due to the high heritability (.72–.74) for this trait, which is opening new opportunities for improving early growth through genomic selection. Selection for early growth could shorten generation time of striped catfish *P. hypophthalmus* which often takes 3–4 years to maintain a breeding cycle in genetic improvement programs. However, selection for tag weight may not capture all genetic variation in body traits at harvest as the genetic correlation ($r_g$) between these two traits is reported to be .5 in this population (Vu et al., 2019b). In Asian seabass *Lates calcarifer*, Khang et al. (2018) also observed a significantly different from one genetic correlation ($r_g$ = .31–.47) for body weights between successive rearing periods from 180 to 556 days post-hatch. Based on the genetic correlation estimates between tag and harvest weights, it is necessary to examine genomic prediction accuracy for harvest weight in this population of striped catfish. Furthermore, there are also no clear advantages regarding the prediction accuracy of ssGBLUP and ssGWAS as compared with PBLUP in our study. Future work should consider enlarging the sample size (in terms of the number of individuals and families genotyped) and number of SNPs to take the advantages of ssGBLUP and ssGWAS models that can capture some measures of Mendelian sampling to improve the estimation of genetic (kinship) matrices for all individuals in the pedigree and hence, improving accuracy of estimated breeding values for tagging weight in this population of striped catfish.

## 4.2 Does omission of the common full-sib effects affect the genomic prediction accuracy?

When the common full-sib families ($c^2$) were excluded from our statistical methods, this resulted in upward biases in the prediction accuracy by 96.5%–130.3% for tagging weight. The overestimation of the prediction accuracy was to a greater extent when PBLUP was used as compared with other methods (i.e., GBLUP and ssGBLUP). To date, no published information is available in aquaculture species to compare with our studies. However, studies in farmed animals suggested that effects of non-additive genetics should be included in mating structures to improve accuracy of genomic prediction and hence, maximizing productivity for dairy farms (Aliloo et al., 2017; Varona et al., 2018). Conventional genetic evaluation systems using pedigree and phenotype data in aquaculture species have also shown that the animal breeding values (EBVs) estimated for growth traits were overestimated, for instance, 10%–56% in giant freshwater prawn *M. rosenbergii* (Phuc et al., 2021) or red tilapia *O.* spp. (Nguyen et al., 2017). Hence, our results are as expected because the $c^2$ estimates were often large for growth traits in aquaculture species where separate rearing of each family was often conducted over a period of 2–3 months until the fish reach a suitable size (e.g., 10–20 g) for physical tagging. The $c^2$ estimates were generally not significant if early communal rearing of all families is practised and DNA markers are used for parentage assignment, as demonstrated in

common carp *C. carpio* (Ninh et al., 2013) or yellowtail kingfish *S. lalandi* (Premachandra et al., 2017). Collectively, due to the high $c^2$ effects on tagging weight and its low to moderate genetic correlation with market (harvest) weight, genomic evaluation models for these traits should account for the common full-sib families and they should be considered as separate traits in genetic improvement programs for striped catfish as well as other aquaculture species.

## 4.3 Can multivariate analysis lessen the upward biases in the prediction accuracy?

Our multivariate analysis of tagging weight in combination with disease resistance trait (survival time) aimed to utilise genetic covariation between the traits and hence improved the predictive power of statistical models used. In addition, when the $c^2$ were omitted, the extent of the overestimation in the prediction accuracy was smaller in the multivariate analysis than univariate models. Studies in animals and plants have reported that multi-trait analysis can improve the prediction accuracy for productivity traits (e.g., milk yield in cattle or grain yield in wheat *T. aestivum L.*) by 0%–28.5% (Sandhu et al., 2021). However, other studies also showed that there are little or no benefits of multivariate vs. single trait analysis (Kemper et al., 2018). To date, studies in aquaculture species performed multi-trait genomic prediction are limited. Results from these studies showed that the accuracies of genomic predictions were not improved for fillet weight and fillet yield in Nile tilapia *O. niloticus* (Joshi et al., 2020) or for survival status and survival time in striped catfish *P. hypophthalmus* (Vu et al., 2021), likely because the high heritability of these two traits and their high genetic correlations; hence, adding one trait did not improve the prediction accuracy of the other. In yellowtail kingfish, Nguyen et al. (2022) also showed that the benefits of multi- vs. univariate analysis depend on statistical methods used and genomic architecture of traits. Hence, molecular dissection of the genomic architecture of traits (e.g., identifying pleotropic loci) can help further understand the impacts of multi-trait analysis on the prediction accuracy for tagging weight and disease resistance examined in this population.

## 4.4 What can imputation help in genomic prediction?

In this study, we found that imputation of missing genotypes has two major benefits. First, it improved the prediction accuracy for tagging weight by 2.1%–12.8%, as compared with when the original (un-imputed) data were used. Second, the imputation reduced the upward biases in the prediction accuracy for tagging weight when the $c^2$ estimates were omitted from our statistical models, mainly because the complete genotypes improved accuracy of estimated breeding values for tagging weight. The benefit of imputation on genomic prediction in aquaculture breeding has been reported in recent studies, such as for disease resistance to photobacteriosis in gilthead sea bream *S. aurata* (Bargelloni et al., 2021), resistance to sea lice in Atlantic salmon *S. salar* (Tsai et al., 2017; Kjetså et al., 2020), growth-related traits Yellowtail kingfish *Seriola lalandi* (Nguyen et al., 2018a) or with simulated data in rainbow trout *O. mykiss* (Dufflocq et al., 2019). In selective breeding programs, imputation can help to reduce costs associated with sequencing. One option is to perform low-density genome sequence (Kriaridou et al., 2020) for a large number of selection candidates and high-density sequence for only parents (Tsai et al., 2017). Then imputation is made to impute from low to high or whole genome sequence. This would help increase selection intensity and thus genetic gain made in selected populations. Furthermore, when more data are accumulated in this population, imputation can increase power of detecting variants for tag weight in genome-wide association studies or fine mapping analysis, integrate multi-studies for meta-analysis of datasets, which are genotyped on different platforms or level of genome coverage. However, also note that the performance of genotype or sequence imputation is affected by many factors, such as reference selection, SNP density, sample size, sequence coverage, minor allele frequency of populations (Chen et al., 2014; Druet et al., 2014; Dufflocq et al., 2019). These factors are fully or partially accounted for in recent software packages that can facilitate the imputation in our breeding program for high growth in striped catfish *P. hypophthalmus*.

## 4.5 Can ssGWAS alleviate the impacts of the $c^2$ omission on the prediction accuracy?

Inclusion of highly significant SNPs in genomic prediction models that included the $c^2$ did not have noticeable impacts on the prediction accuracy for tagging weight. This is likely due to the limited size of the significant SNPs obtained from genotyping by sequencing (GBS) platform but our observation here is consistent with previous findings for disease resistance traits in the same population of striped catfish *P. hypophthalmus* (Vu et al., 2021). In studies where the $c^2$ estimates were not included, Luo et al. (2021) also found there were no advantages of pre-selected SNPs in genomic prediction models using ssGBLUP, WssGBLUP and BayesB for resistance to *Edwardsiella tarda* that causes acute symptoms with ascites in Japanese flounder (*Paralichthys olivaceus*). However, other studies, which used prioritised variants from GWAS, reported there was an

improvement in the prediction accuracy by 1.2%–13.3% for growth-related traits under chronic thermal stress in rainbow trout *O. mykiss* (Yoshida and Yáñez, 2021) or disease resistance traits in whiteleg shrimp *L. vannamei*, Atlantic salmon *S. salar* and gilthead sea bream *S. aurata* (Luo et al., 2021). In addition, the variant (or marker) effects can be weighed to improve the prediction accuracy as demonstrated in our recent study for disease traits (Vu et al., 2021) or for production traits in dairy cattle (Xiang et al., 2021).

## 5 Concluding remarks

The prediction accuracy for tagging weight using BLUP-family methods was moderate to high. The omission of the common full-sib families resulted in upward biases in the predictive performance across statistical models used. Imputation of missing values alleviated the impacts of the common full-sib families on the prediction accuracy. As compared with single trait analysis, multivariate model slightly improved the prediction accuracy when the $c^2$ effects were excluded from our analyses. A combined ssGWAS with ssGBLUP did not sacrifice the prediction accuracy, regardless of the $c^2$. Our results suggest that genomic selection for early growth traits should include the $c^2$ in statistical models to investigate any possible changes in selection accuracy and selection response. Future study should increase the number of genotyped individuals and/or consider alternative genotyping platforms (e.g., whole genome sequencing) as well as use different mating structures (e.g., using full or partial factorial design) to enable the separation of the dominance from common full-sib effects in order to improve accuracy of genomic prediction for tagging weight and commercial traits of economic importance in this striped catfish population.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

## Ethics statement

The animal study was reviewed and approved by Research Institute for Aquauclture No. 2, Vietnam.

## Author contributions

NV, TP, NN, and NV designed, conducted the study, collected and analysed the data and wrote and approved the manuscript for submission.

## Funding

Ministry of Agriculture and Rural Development of Vietnam and University of the Sunshine Coast (USC) in Queensland, Australia provided financial support for this project.

## Acknowledgments

We expressed our deep gratitude to the research team at National Breeding Center for Southern Freshwater Aquaculture of RIA2 for their collections of phenotypes and DNA samples used in the study. We also extend our sincere thanks to Phuong Thi Kim Oanh for her early involvement in the disease challenge test experiment.

## Conflict of interest

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1081246/full#supplementary-material

## References

Aguilar, I., Legarra, A., Cardoso, F., Masuda, Y., Lourenco, D., and Misztal, I. (2019). Frequentist p-values for large-scale-single step genome-wide association, with an application to birth weight in American Angus cattle. *Genet. Sel. Evol.* 51 (1), 28–8. doi:10.1186/s12711-019-0469-3

Al-Tobasei, R., Ali, A., Garcia, A. L. S., Lourenco, D., Leeds, T., and Salem, M. (2021). Genomic predictions for fillet yield and firmness in rainbow trout using reduced-density SNP panels. *BMC Genomics* 22 (1), 92. doi:10.1186/s12864-021-07404-9

Aliloo, H., Pryce, J., González-Recio, O., Cocks, B., Goddard, M., and Hayes, B. (2017). Including nonadditive genetic effects in mating programs to maximize dairy farm profitability. *J. Dairy Sci.* 100 (2), 1203–1222. doi:10.3168/jds.2016-11261

Allal, F., and Nguyen, N. H. (2022). "Genomic selection in aquaculture species," in *Complex trait prediction* (Germany: Springer), 469–491.

Bargelloni, L., Tassiello, O., Babbucci, M., Ferraresso, S., Franch, R., Montanucci, L., et al. (2021). Data imputation and machine learning improve association analysis and genomic prediction for resistance to fish photobacteriosis in the gilthead sea bream. *Aquac. Rep.* 20, 100661. doi:10.1016/j.aqrep.2021.100661

Bosworth, B., Waldbieser, G., Garcia, A., Tsuruta, S., and Lourenco, D. (2020). Heritability and response to selection for carcass weight and growth in the Delta Select strain of channel catfish, I*ctalurus punctatus*. *Aquaculture* 515, 734507. doi:10.1016/j.aquaculture.2019.734507

Bui, T. M., Phan, L. T., Ingram, B. A., Nguyen, T. T., Gooley, G. J., Nguyen, H. V., et al. (2010). Seed production practices of striped catfish, *Pangasianodon hypophthalmus* in the Mekong Delta region. *Vietnam. Aquac.* 306 (1-4), 92–100. doi:10.1016/j.aquaculture.2010.06.016

Chen, L., Li, C., Sargolzaei, M., and Schenkel, F. (2014). Impact of genotype imputation on the performance of GBLUP and Bayesian methods for genomic prediction. *PLoS One* 9 (7), e101544. doi:10.1371/journal.pone.0101544

Druet, T., Macleod, I., and Hayes, B. (2014). Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* 112 (1), 39–47. doi:10.1038/hdy.2013.13

Dufflocq, P., Pérez-Enciso, M., Lhorente, J. P., and Yáñez, J. M. (2019). Accuracy of genomic predictions using different imputation error rates in aquaculture breeding programs: A simulation study. *Aquaculture* 503, 225–230. doi:10.1016/j.aquaculture.2018.12.061

Gianola, D., and Fernando, R. L. (2020). A multiple-trait bayesian lasso for genome-enabled analysis and prediction of complex traits. *Genetics* 214 (2), 305–331. doi:10.1534/genetics.119.302934

Gutierrez, A. P., Matika, O., Bean, T. P., and Houston, R. D. J. F. i. g. (2018). Genomic selection for growth traits in pacific oyster (*Crassostrea gigas*): Potential of low-density marker panels for breeding value prediction. *Front. Genet.* 9, 391. doi:10.3389/fgene.2018.00391

Hamzah, A., Thoa, N. P., and Nguyen, N. H. (2017). Genetic analysis of a red tilapia (Oreochromis spp.) population undergoing three generations of selection for increased body weight at harvest. *J. Appl. Genet.* 58 (4), 509–519. doi:10.1007/s13353-017-0411-8

Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., et al. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nat. Rev. Genet.* 21 (7), 389–409. doi:10.1038/s41576-020-0227-y

Joshi, R., Skaarud, A., de Vera, M., Alvarez, A. T., and Ødegård, J. (2020). Genomic prediction for commercial traits using univariate and multivariate approaches in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 516, 734641. doi:10.1016/j.aquaculture.2019.734641

Kemper, K. E., Bowman, P. J., Hayes, B. J., Visscher, P. M., and Goddard, M. E. (2018). A multi-trait Bayesian method for mapping QTL and genomic prediction. *Genet. Sel. Evol.* 50 (1), 10–13. doi:10.1186/s12711-018-0377-y

Khang, P. V., Phuong, T. H., Dat, N. K., Knibb, W., and Nguyen, N. H. (2018). An 8-year breeding program for Asian seabass Lates calcarifer: Genetic evaluation, experiences, and challenges. *Front. Genet.* 9, 191. doi:10.3389/fgene.2018.00191

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., et al. (2012). "Diversity arrays technology: A generic genome profiling technology on open platforms," in *Data production and analysis in population genomics* (Germany: Springer), 67–89.

Kjetså, M., Ødegård, J., and Meuwissen, T. (2020). Accuracy of genomic prediction of host resistance to salmon lice in Atlantic salmon (*Salmo salar*) using imputed high-density genotypes. *Aquaculture* 526, 735415. doi:10.1016/j.aquaculture.2020.735415

Kriaridou, C., Tsairidou, S., Houston, R. D., and Robledo, D. (2020). Genomic prediction using low density marker panels in aquaculture: Performance across species, traits, and genotyping platforms. *Front. Genet.* 11, 124. doi:10.3389/fgene.2020.00124

Luan, S., Yang, G., Wang, J., Luo, K., Zhang, Y., Gao, Q., et al. (2012). Genetic parameters and response to selection for harvest body weight of the giant freshwater prawn *Macrobrachium rosenbergii*. *Aquaculture* 362, 88–96. doi:10.1016/j.aquaculture.2012.05.011

Luo, Z., Yu, Y., Xiang, J., and Li, F. (2021). Genomic selection using a subset of SNPs identified by genome-wide association analysis for disease resistance traits in aquaculture species. *Aquaculture* 539, 736620. doi:10.1016/j.aquaculture.2021.736620

Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., Legarra, A., and Vitezica, Z. (2018). *Manual for BLUPF90 family of programs*. United States: University of Georgia.

Montesinos-López, O. A., Montesinos-López, A., Pérez-Rodríguez, P., Barrón-López, J. A., Martini, J. W., Fajardo-Flores, S. B., et al. (2021). A review of deep learning applications for genomic selection. *BMC Genomics* 22 (1), 19–23. doi:10.1186/s12864-020-07319-x

Montesinos-López, O. A., Montesinos-López, J. C., Singh, P., Lozano-Ramirez, N., Barrón-López, A., Montesinos-López, A., et al. (2020). A multivariate Poisson deep learning model for genomic prediction of count data. *G3* 10 (11), 4177–4190. doi:10.1534/g3.120.401631

Nguyen, N. H. (2021). A systematic review and meta-analysis of genetic parameters for complex quantitative traits in aquatic animal species. bioRxiv. doi:10.1101/2021.1105.1120.445048

Nguyen, N. H., Hamzah, A., and Ngo, T. P. (2017). Effects of genotype by environment interaction on genetic gain and genetic parameter estimates in red Tilapia (*Oreochromis* spp.). *Front. Genet.* 8, 82. doi:10.3389/fgene.2017.00082

Nguyen, N. H., Ninh, N. H., and Hung, N. H. (2020a). Evaluation of two genetic lines of Pacific White leg shrimp *Liptopenaeus vannamei* selected in tank and pond environments. *Aquaculture* 516, 734522. doi:10.1016/j.aquaculture.2019.734522

Nguyen, N. H., Phuthaworn, C., and Knibb, W. (2020b). Genomic prediction for disease resistance to Hepatopancreatic parvovirus and growth, carcass and quality traits in Banana shrimp *Fenneropenaeus merguiensis*. *Genomics* 112 (2), 2021–2027. doi:10.1016/j.ygeno.2019.11.014

Nguyen, N. H., Premachandra, H., Kilian, A., and Knibb, W. (2018a). Genomic prediction using DArT-Seq technology for yellowtail kingfish *Seriola lalandi*. *BMC Genomics* 19 (1), 107. doi:10.1186/s12864-018-4493-4

Nguyen, N. H., Rastas, P. M., Premachandra, H., and Knibb, W. (2018b). First high-density linkage map and single nucleotide polymorphisms significantly associated with traits of economic importance in Yellowtail Kingfish *Seriola lalandi*. *Front. Genet.* 9, 127. doi:10.3389/fgene.2018.00127

Nguyen, N. H., Vu, N. T., Patil, S. S., and Sandhu, K. S. (2022). Multivariate genomic prediction for commercial traits of economic importance in Banana shrimp Fenneropenaeus merguiensis. *Aquaculture* 555, 738229. doi:10.1016/j.aquaculture.2022.738229

Ninh, N. H., Ponzoni, R. W., Nguyen, N. H., Woolliams, J. A., Taggart, J. B., McAndrew, B. J., et al. (2013). A comparison of communal and separate rearing of families in selective breeding of common carp (*Cyprinus carpio*): Responses to selection. *Aquaculture* 408409 (0), 152–159. doi:10.1016/j.aquaculture.2013.06.005

Okeke, U. G., Akdemir, D., Rabbi, I., Kulakow, P., and Jannink, J.-L. (2017). Accuracies of univariate and multivariate genomic prediction models in African cassava. *Genet. Sel. Evol.* 49 (1), 88. doi:10.1186/s12711-017-0361-y

Palaiokostas, C., Kocour, M., Prchal, M., and Houston, R. D. (2018). Accuracy of genomic evaluations of juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by sequencing. *Front. Genet.* 9 (82), 82. doi:10.3389/fgene.2018.00082

Phuc, T. H., Vu, N. T., Nga, N. T. K., Ky, N. T., and Nguyen, N. H. (2021). Assessment of a long-term selective breeding program for giant freshwater prawn Macrobrachium rosenbergii since 2007. *Aquaculture* 541, 736745. doi:10.1016/j.aquaculture.2021.736745

Premachandra, H. K. A., Nguyen, N. H., Miller, A., D'Antignana, T., and Knibb, W. (2017). Genetic parameter estimates for growth and non-growth traits and comparison of growth performance in sea cages vs land tanks for yellowtail kingfish *Seriola lalandi*. *Aquaculture* 479, 169–175. doi:10.1016/j.aquaculture.2017.05.043

Samorn, P. (2007). Effects of equilibration times on the fertilization rate of cryopreserved striped catfish, pangasius hypophthalmus (sauvage, 1878) sperm. *Aquac. Res.* 34, 887–893. doi:10.1046/j.1365-2109.2003.00897.x

Sandhu, K. S., Patil, S. S., Pumphrey, M. O., and Carter, A. H. (2021). Multi-trait machine and deep learning models for genomic selection using spectral information in a wheat breeding program. *Plant Genome* 14 (3), e20119. doi:10.1002/tpg2.20119

Sang, N. V., Luan, N. T., Hao, N. V., Nhien, T. V., Vu, N. T., and Nguyen, N. H. (2020a). Genotype by environment interaction for survival and harvest body weight between recirculating tank system and pond culture in *Penaeus monodon*. *Aquaculture* 525, 735278. doi:10.1016/j.aquaculture.2020.735278

Sang, V. V., Knibb, W., Nguyen, N. T. H., Vu, I. V., O'Connor, W., Dove, M., et al. (2020b). First breeding program of the Portuguese oyster *Crassostrea angulata* demonstrated significant selection response in traits of economic importance. *Aquaculture* 518, 734664. doi:10.1016/j.aquaculture.2019.734664

Sukhavachana, S., Poompuang, S., Onming, S., and Luengnaruemitchai, A. (2019). Heritability estimates and selection response for resistance to Streptococcus agalactiae in red tilapia Oreochromis spp. *Aquaculture* 502, 384–390. doi:10.1016/j.aquaculture.2018.12.075

Tsai, H.-Y., Matika, O., Edwards, S. M., Antolín–Sánchez, R., Hamilton, A., Guy, D. R., et al. (2017). Genotype imputation to improve the cost-efficiency of genomic selection in farmed Atlantic salmon. *G3* 7 (4), 1377–1383. doi:10.1534/g3.117.040717

Tsuruta, S., and Misztal, I. (2006). "THRGIBBS1F90 for estimation of variance components with threshold and linear models," in Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13–18 August 2006.

van den Berg, S., Calus, M. P., and Wientjes, Y. (2015). Across population genomic prediction scenarios in which Bayesian variable selection outperforms GBLUP. *BMC Genet.* 16 (1), 146. doi:10.1186/s12863-015-0305-x

Van Sang, N., Klemetsdal, G., Ødegård, J., and Gjøen, H. M. (2012). Genetic parameters of economically important traits recorded at a given age in striped catfish (*Pangasianodon hypophthalmus*). *Aquaculture* 344, 82–89. doi:10.1016/j.aquaculture.2012.03.013

Varona, L., Legarra, A., Toro, M. A., and Vitezica, Z. G. (2018). Non-additive effects in genomic selection. *Front. Genet.* 9, 78. doi:10.3389/fgene.2018.00078

Vu, N. T., Ha, T. T. T., Thuy, V. T. B., Trang, V. T., and Nguyen, N. H. (2020). Population genomic analyses of wild and farmed striped catfish *Pangasianodon hypophthalmus* in the lower mekong river. *J. Mar. Sci. Eng.* 8 (6), 471. doi:10.3390/jmse8060471

Vu, N. T., Sang, N. V., Trong, T. Q., Duy, N. H., Dang, N. T., and Nguyen, N. H. (2019a). Breeding for improved resistance to Edwardsiella ictaluri in striped catfish (Pangasianodon hypophthalmus): Quantitative genetic parameters. *J. Fish. Dis.* 42 (10), 1409–1417. doi:10.1111/jfd.13067

Vu, N. T., Van Sang, N., Phuc, T. H., Vuong, N. T., and Nguyen, N. H. (2019b). Genetic evaluation of a 15-year selection program for high growth in striped catfish Pangasianodon hypophthalmus. *Aquaculture* 509, 221–226. doi:10.1016/j.aquaculture.2019.05.034

Vu, T. N., Tran, P. H., Kim, O. T. P., Van Nguyen, S., Trinh, T. T., and Nguyen, N. H. (2021). Accuracies of genomic predictions for disease resistance of striped catfish to *Edwardsiella ictaluri* using artificial intelligence algorithms. *G3* 12 (1), jkab361. doi:10.1093/g3journal/jkab361

Whalen, A., Gorjanc, G., and Hickey, J. M. (2020). AlphaFamImpute: High-accuracy imputation in full-sib families from genotype-by-sequencing data. *Bioinformatics* 36 (15), 4369–4371. doi:10.1093/bioinformatics/btaa499

Xiang, R., MacLeod, I. M., Daetwyler, H. D., de Jong, G., O'Connor, E., Schrooten, C., et al. (2021). Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations. *Nat. Commun.* 12 (1), 860–913. doi:10.1038/s41467-021-21001-0

Yin, L., Zhang, H., Zhou, X., Yuan, X., Zhao, S., Li, X., et al. (2020). Kaml: Improving genomic prediction accuracy of complex traits using machine learning determined parameters. *Genome Biol.* 21 (1), 146–222. doi:10.1186/s13059-020-02052-w

Yoshida, G. M., and Yáñez, J. M. (2021). Increased accuracy of genomic predictions for growth under chronic thermal stress in rainbow trout by prioritizing variants from GWAS using imputed sequence data. *Evol. Appl.* 15, 537–552. doi:10.1111/eva.13240

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fgene.2022.1047436



OPEN ACCESS

EDITED BY
Md Samsul Alam,
Bangladesh Agricultural University, Bangladesh

REVIEWED BY
Asaduzzaman Md,
Chattogram Veterinary and Animal Sciences University, Bangladesh
Vindhya Mohindra,
National Bureau of Fish Genetic Resources (ICAR), India

*CORRESPONDENCE
Imtiaz Ahmed,
imtiazamu1@yahoo.com
Syed Mudasir Ahmad,
mudasirbio@gmail.com

SPECIALTY SECTION
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

RECEIVED 18 September 2022
ACCEPTED 29 November 2022
PUBLISHED 04 January 2023

CITATION
Awas M, Ahmed I, Ahmad SM, Al-Anazi KM, Farah MA and Bhat B (2023), Integrative approach for validation of six important fish species inhabiting River Poonch of north-west Himalayan region (India). *Front. Genet.* 13:1047436. doi: 10.3389/fgene.2022.1047436



# Integrative approach for validation of six important fish species inhabiting River Poonch of north-west Himalayan region (India)

Mohd Awas[1], Imtiaz Ahmed[1]*, Syed Mudasir Ahmad[2]*, Khalid Mashay Al-Anazi[3], Mohammad Abul Farah[3] and Basharat Bhat[2]

[1]Fish Nutrition Laboratory, Department of Zoology, University of Kashmir, Srinagar, India, [2]Division of Animal Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, Sher-e-Kashmir University of Agricultural Sciences and Technology, Srinagar, India, [3]Department of Zoology, College of Science, King Saud University, Riyadh, Saudi Arabia

Traditionally, species of fish are identified based on morphological characteristics. Although these taxonomic descriptions are essential, there are cases where the morphological characters distinguishing these species show marginal differences. For instance, in the Poonch River in the Himalayas, there are 21 species, out of which some are morphologically similar, and the taxonomic distinction between these species is unclear. Therefore, in this study, we used sequences from two mitochondrial genes, Cytochrome b (*Cyt b*) and a larger ribosomal subunit (16S rRNA), as well as the morphological analysis to address any taxonomic ambiguities among the six fish species. Maximum Likelihood results revealed that all the species were clustered according to their families and genera. The phenotypic analysis also supported this statement, as all the species of different genera like *Schizothorax*, *Tor*, *Garra*, *Traqilabeo*, and *Glyptothorax* are grouped in their particular cluster, it shows that species of a separate class share a mutual morphological characteristic. While genetic analyses of these species suggest nucleotide diversity (p) and haplotype diversity, with Hd values as 0.644 for *Cyt b* and 0.899 for 16S rRNA, confirming the rich genetic diversity in the river. Overall, we recommend that the integrative approach in delimiting the fish species is more effective than the individual one and can be used to rapidly diagnose a species and understand the evolutionary relationship between the species.

KEYWORDS

**fish diversity, Poonch River, *Cyt b*, 16S rRNA, validation, identification**

## Introduction

Over the last few decades, fish consumption has gradually become an important source of protein (FAO, 2000). The importance of fish as a source of human nutrition cannot be overstated because it contains high-quality protein, minerals and lipids high in omega-3 unsaturated fatty acids (Bavinck and Johnson, 2008), thus considered one of the most valuable components of the human diet. Besides, fisheries are important sources of revenue for many communities (Rafique and Khan, 2012; Ghouri et al., 2020). Over 33,000 species of fish are known, which inhabit nearly all major aquatic habitat types and contribute a wide range of environmental functions (Martinez et al., 2018). Around 2,500 fish species harbour Indian water and 930 are freshwater fish species (Jayaram, 2010). In India, 788 fish species come under freshwater fishes (Gopi, 2014). At the same time, the Himalayan region contributes about 17% of fish documented from the mountain ecosystem (Ghosh, 1997). In the present study, we have focused on the Poonch River of the north-west Himalayas, which originates from the foothill of Pir-panjal (Ahmed et al., 2019; Awas et al., 2020). This region is known for harboring diverse aquatic fauna, including 21 fish species, out of which six are considered economically important species (Hussain and Dutta, 2016).

Dutta (2003) reported 40 fish species from the Poonch River, including several species having sub-species under the same genus as *Schizothorax*, *Tor*, *Garra*, *Crossocheilus*, and *Glyptothorax*, etc. Such a considerable variation between these statements might be due to species ambiguity. Moreover, several findings suggest a vide ambiguity in *Schizothorax* and *Tor* genus (He and Chen, 2006; Yang et al., 2012; Ahmad et al., 2014; Dwivedi et al., 2020; Jaafar et al., 2021). However, in the region, these species under two genera are mainly considered as *Schizothorax richardsonii* (Luss) and *Tor putitora* (Chidak) due to their similarities in morphological traits. While Hussain and Dutta (2016) reported only *S. richardsonii* and *T. Putitora* belongs to the two genera from the Poonch River. Similarly, *Garra gotyla*, *Garra lamta*, *Tariqilabeo latius* and *Tariqilabeo diplocheius* have been reported by Dutta (2003) under four different species of two genera. But, due to the similarities in external morphological characteristics, there is a huge ambiguity among these species also, as Hussain and Dutta (2016) reported three species of two genera from this region.

To avoid mislabeling in fish markets, fish must be identified authentically and reliably. Morphological and morphometric characterization of fish species is one of the most often used strategies for fish authenticity (Bottero and Dalmasso, 2011). But limitations of classical taxonomy led to ambiguous categorization of fishes in the event of very close similarities between species and ultimately give an inaccurate picture of the ichthyofaunal diversity of the area (Gonzalez et al., 2013; Xu et al., 2018). Nowadays specific molecular techniques have been employed to overcome the constraints of morphological-based identification systems in fish (Hebert et al., 2003; Griffiths et al., 2013; Keskin and Atar, 2013; Zhuang et al., 2013; Knebelsberger et al., 2014; Diaz et al., 2016; Shen et al., 2016; Ghouri et al., 2020).

For fish identification, many DNA biomarkers have also been used. DNA barcoding has widely been used in different sectors, including fish authentication, labeling, biodiversity, conservation and biological studies (Pollack et al., 2018). The Cytochrome b (*Cyt b*) gene, which has ubiquitous uses in taxonomic and ecological domains, is one of the most well-known and targeted DNA markers. It is often used to study the phylogenetic relationships among organisms due to its small size and high nucleotide substitution rate (Xiao et al., 2001; Kumar et al., 2011; Ahmad et al., 2014; Cornetti et al., 2019). In addition, 16S ribosomal RNA has gained popularity due to its extremely high degree of conservation and relatively slow evolution rate (Hebert et al., 2003; Chen et al., 2018). Several past studies have shown that the 16S rRNA gene sequence is also helpful for species identification (Kochzius et al., 2010; Jaafar et al., 2021).

Most fish species inhabiting the Poonch River belong to the family Cyprinidae, represented by a group of species with highly similar external morphological characters. The survival of *Glyptothorax kashmirensis* (Kashmiri catfish) and *T. Putitora* (Mahaseer) exclusively depends on this water source. However, in 2010 Poonch River gained much more attention and was designated as a national park (Brown et al., 2019), but the lack of study based on the fish fauna is still marginal. Therefore, the fish species that have been recorded in the Poonch River; six have not been clearly classified by previous studies based on the analysis of morphological characters. To identify and classify these species unambiguously, we conducted this study using an integrative approach based on analyzing several morphological parameters and molecular research with *Cyt b* and 16S rRNA as genetic markers. The morphometric parameters used during the study were according to Froese and Pauly (2007). The different methods like PCA, CVA, RDA, etc. were used to get a more robust result; moreover, Jayaram (2010) standard keys were used for species determination. Besides, we also determine the efficacy of the selected genes in identifying the freshwater fish fauna. This study could be useful in examining DNA sequences, which might serve as a primary reference for future region studies.

## Materials and methods

### Fish sample collection

In the present study, 478 specimens were sampled from 2018 to 2020 from six different sites in Poonch River with the help of native professional fishermen by using different fishing gears like cast net, gillnet and hooks, while the details about the number of samples collected, sampling sites and geo-graphical

**TABLE 1** Sampling sites, their locations with latitude and longitude, sample size and size range of six fish species used in the present study with GenBank accession number.

| Species | Standard length (SL) range (cm) | Sample size | | Sampling sites | Location | Cyt b | 16S rRNA | IUCN status |
|---|---|---|---|---|---|---|---|---|
| | | Traditional morphometrics | Molecular analysis | | | | | |
| *G. gotyla* | 12.2–18.5 | 79 | 9 | Madainy | 33°40′42.57″N–74°15′7.35″E | MW191577, MW191578 | MW148586, MW148585 | LC |
| | | | | Kalai | 33°44′20.57″N–74°1′36.04″E | | | |
| | | | | Bufflaiz | 33°6′13.05″N–74°35′9.28″E | | | |
| | | | | Mankote | 33°37′59.14″ N–74°3′54.02″ E | | | |
| *T. putitora* | 17.1–28.8 | 77 | 8 | Mendhar | 33°36′24.07″N–74°8′36.74″E | MW191579, MW191580, MW191581 | MW148589 | EN |
| | | | | Sher-e-Kashmir bridge | 33°7′58.33″N–74°9′19.44″E | | | |
| | | | | Mankote | 33°37′59.14″N–74°3′54.02″E | | | |
| | | | | Mendhar | 33°36′24.07″N–74°8′36.74″E | | | |
| | | | | Sher-e-Kashmir bridge | 33°7′58.33″N–74°9′19.44″E | | | |
| | | | | Mankote | 33°37′59.14″N–74°3′54.02″E | | | |
| | | | | Kalai | 33°44′20.57″N–74°1′36.04″E | | | |
| *S. richardsonii* | 10.32–31.76 | 79 | 8 | Bufflaiz | 33°6′13.05″N–74°35′9.28″E | MW191586 | MW148587 | VU |
| | | | | Mendhar | 33°36′24.07″N–74°8′36.74″E | | | |
| | | | | Kalai | 33°44′20.57″N–74°1′36.04″E | | | |
| | | | | Sher-e-Kashmir bridge | 33°7′58.33″N–74°9′19.44″E | | | |
| | | | | Madainy | 33°40′42.57″N–74°15′7.35″E | | | |
| *S. plagiostomus* | 14.2–28.9 | 78 | 7 | Madainy | 33°40′42.57″N–74°15′7.35″E | | MW148588 | NA |
| | | | | Sher-e-Kashmir bridge | 33°7′58.33″N–74°9′19.44″E | | | |
| | | | | Kalai | 33°44′20.57″N–74°1′36.04″E | | | |
| | | | | Bufflaiz | 33°6′13.05″N–74°35′9.28″E | | | |

(Continued on following page)





10.3389/fgene.2022.1047436

**TABLE 1** (*Continued*) Sampling sites, their locations with latitude and longitude, sample size and size range of six fish species used in the present study with GenBank accession number.

| Species | Standard length (SL) range (cm) | Sample size | | Sampling sites | Location | Cyt b | 16S rRNA | IUCN status |
|---|---|---|---|---|---|---|---|---|
| | | Traditional morphometrics | Molecular analysis | | | | | |
| *T. latius* | 11.1–16.2 | 79 | 8 | Madainy | 33°40′42.57″N–74°15′7.35″E | MW191585 | MW148582, MW148583, MW148584 | LC |
| | | | | Mankote | 33°37′59.14″N–74°3′54.02″E | | | |
| | | | | Kalai | 33°44′20.57″N–74°1′36.04″E | | | |
| | | | | Bufflaiz | 33°6′13.05″N–74°35′9.28″E | | | |
| *G. kashmirinsis* | 6.9–11.2 | 37 | 2 | Bufflaiz | 33°6′13.05″N–74°35′9.28″E | | MW148582 | CR |
| | | | | Madainy | 33°40′42.57″N–74°15′7.35″E | | | |
| | | | | Sher-e-Kashmir bridge | 33°7′58.33″N–74°9′19.44″E | | | |

coordinates of locations for fish species investigated in the current study is provided in Table 1 and Figure 1, and the representative images of each species has shown in Supplementary Figure S1. All these specimens belonging to six species were carried out in an ice box at the Wet Laboratory, Department of Zoology, the University of Kashmir, for further study.

## Identification

The collected fish species were first identified using standard taxonomic keys and related books (Talwar and Jhingran, 1991; Vishwanath et al., 2007; Jayaram, 2010). After proper identification, representative specimens were preserved in 10% formalin in the Fish Nutrition Research Laboratory, Department of Zoology, University of Kashmir Srinagar, India, for museum records.

## Morphometric and meristic controlling elements

Different morphometric and meristic controlling elements were also taken across the fish body by following different methods reported in the past (Menon, 1964; Gonzales and Sartori, 2002; Mina et al., 2005; Froese and Pauly, 2007). The measurements were recorded using a 0.01 cm digital vernier caliper from left to right. Overall, 25 phenotypic characters were analyzed. Counts and measurements were taken as far as possible on the left side of fish specimens following standard methods for cyprinid taxonomy (Pethiyagoda, and Kottelat, 1995) with some additional modifications. The morphometric characters measured during the present study (Supplementary Figure S2) included total length (TL), standard length (SL), forked length (FL), pre-pectoral length (PpL), pre-pelvic length (ppL), pre-dorsal length (PDL), pre-anal length (PAL), pectoral fin length (PFL), pectoral fin height (PFH), pelvic fin length (pFL), pelvic fin height (pFH), dorsal fin length (DFL), dorsal fin height (DFH), anal fin length (AFL), anal fin height (AFH), caudal fin length (CFL), caudal fin height (CFH), head length (HL), snout length (SL), eye diameter (E-dia), pre-orbital length (PorbL), minimum body depth (MinBD), maximum body depth (MaxBD). While meristic characters were studied, including dorsal fin rays (DFR), pectoral fin rays (PFR), pelvic fin rays (pFR), anal fin rays (AFR), and caudal fin rays (CFR). Meristic characters were counted twice by the same observer. In the current study, it was hypothesized that a working relationship exists between various growth-related morphometric parameters. The correlation coefficient "r" was determined among all the characteristics.


FIGURE 1
Map showing six sampling sites of River Pooch of north-west Himalaya India: Site A, Bufflaiz, B, Madainy, C, Kalai, D, Sher e Kashmir, E, Mendhar, F, Mankote.

## Traditional morphometric analysis

Each of the fourteen morphometric characters was divided by standard length and the remaining six characters were divided by head length in order to eliminate the size effect (correlation < .5 for all variables). All the morphometric values were log-transformed before analysis in R software (R Development Core Team, 2020). The regression analysis and the resulting figures were plotted by using "psych" (Wickham, 2016; Revelle, 2020).

## Statistical analysis

All statistical analyses were carried out using R Statistical Software (R Development Core Team, 2020) to perform all morphometric variation analyses among different fish species. We calculated the descriptive statistics for each morphometric parameter, particularly mean standard deviation, using the "dplyr" package (Wickham, et al., 2022). Prior to conducting statistical analysis, we used the Sharpio–Wilk normality test to investigate the normality assumption of the dataset. In order to determine if there were significant differences in morphometric characters, a univariate analysis of variance (ANOVA) was performed. In order to decrease the amount of sample redundancy, a principal component analysis (PCA) was conducted (Johnson and Wichern 1998). Before conducting PCA, the morphometric measurements were subjected to scaling (i.e., standardization) to make the analysis comparable. Moreover, we also evaluated the correlation of studied phenotypic characters with principal components analysis. In general, the variables which are correlated with PC1 or dimension (Dim, 1) and PC2 or Dim, 2 are considered as crucial in order to explain the existing variability among the data set (Kassambara, 2017). The PCA analysis and the resulting figures were obtained using "FactoMinerR" (Le et al., 2008) and "factoextra" (Kassambra and Mundt, 2020) packages. In addition to PCA, multiple variables were constructed and further analytical procedures, such as canonical variate analysis (CVA) and redundancy analysis (RDA), were conducted (Mardia et al., 1979; Ter Braak and Verdonschot, 1995). Moreover, the sample size was taken based on previous work

on PCA and CVA (Kocovsky et al., 2009). The contributions of the variable to PCA were investigated to decide which character contributes more to differentiate the species. For this purpose, the Kaiser–Meyer–Olkin value (0–1) was used to check the suitability for PCA (Kaiser, 1974). Various multivariate analyses, such as PCA, CVA, RDA, and correlations, were performed in R. 4.0.2. However, the dendrogram was constructed to address the dissimilarity between species and the number of clusters in the data set (Charrad et al., 2014).

In order to avoid the effect of sample size, the data were M-transformed by employing the formula given below (Poulet et al., 2005).

$$M\,\text{trans} = \log M - b\,(\log SL - \log SL\,\text{mean})$$

where M–trans is the transformed measurement, M is the original measurement, b is the within-group slope regression of the log standard length, SL is the standard length of the fish species. To assess the effectiveness of size adjustment transformations, the relevance of the relationship between transformed variables and standard length was taken into consideration. The correlation analysis between the modified variables and the fish's standard length was determined to evaluate whether the transformed data successfully removed the size effect or not (Khan et al., 2012).

## Molecular analysis

For molecular analysis, all six species having morphological ambiguities were used to study the diversity and phylogenetic relationship. For barcoding purposes, 2–3 specimens of representative species were preserved in 95% alcohol (Roul et al., 2021; Tsoupas et al., 2022), as the *G. kashmirensis* is one of the protected species; therefore, a small part of the fin was preserved for DNA analysis before releasing the surviving fish back into the water. For further analysis, the samples were transferred to the Genomics Laboratory at Sher-e-Kashmir University of Agricultural Science and Technology, Jammu and Kashmir, India.

## DNA extraction and PCR amplification

Using a sterile razor blade, we extracted DNA from a portion of muscle tissue from the caudal peduncle. The muscle sample was placed in Eppendorf tubes, rinsed with 1 ml phosphate-buffered saline (PBS) in order to remove all contamination, and then the sample was frozen. After discarding (PBS), total genomic DNA was extracted using DNeasy Blood and Tissue Kit (QIAGEN 69504) by following the instruction given by the company. After that, the product was quantified with the help of NanoDrop (Thermo Scientific, Wilmington, NC, United States). Extracted genomic DNA was amplified for the above two genes using universal primers. The primers against 16S rRNA forward (5′ GCC TGT TTA ACA AAA ACA T 3′) and reverse (5′ CCG GTC TGA ACT CAG ATC ATG T3′) (Simon et al., 1991) and Cyt b forward (5′ GTT TGA TCC CGT TTC CTG TA 3′) and reverse (5′ AAT GAC TTG AAG AAC CACCGT 3′) were used for amplification (Briolay et al., 1998). Afterwards, a PCR reaction mixture was performed as per the protocol mentioned in the kit. The following steps were taken for PCR reaction. Initial denaturation for 3 min at 95°C, followed by 34 cycles of 95°C for 30 s (16S) after annealing at 48°C and 50°C for (*Cyt b*), while extension temperature of 72°C for 1 min and final extension of 72°C for 3 min. PCR products were visualized on 1% agarose gel. Finally, the product was purified using a PCR purification kit (Invitrogen, United States) using the company's protocol. Biospectrometer measured the purified PCR products (Thermo Scientific, United States).

## DNA sequencing and data analysis

PCR amplified products were sequenced at genomic Xcelris Labs Limited Ist Floor, Sydney House, Old Premchand Nagar Road, Opp. Satyagrah Chhavani, Bodakdev, Ahmedabad—380015 Gujarat, India. The specific primers successfully amplified the *Cyt b* and 16S rRNA genes of the approximate size of 543 and 519 bp, respectively. The data was aligned with the help of MUSCLE v3.5 (Edgar, 2004) and the ClustalW algorithm (50), whereas the bio-editing tool was done using MacClade v4.06 (Maddison and Maddison, 2003). All the ambivalent bases were removed with the help of ABI chromatograms and the rest of the sequences were submitted to GenBank to get accession numbers. The phylogenetic trees were constructed with the help of PhyML algorithm software using the Maximum likelihood method (Felsenstein, 1981), with jModelTest (Posada, 2008; Darriba et al., 2012) and applied the bootstrap technique set to 100 replicates for testing the trees (Guindon et al., 2009). *Barbus barbus* was used as the out-group. Automatic Barcode Gap Discovery (ABGD) clustering was carried out using Kimura evolutionary model available at (https://bioinfo.mnhn.fr/abi/public/abgd/abgdweb.html) by applying the following parameter: Pmin = .001, Pmax = .1; steps 20; Nb bins = 20; X (gap width) = .75 and Nb bins (for distance distribution) = 20 (Puillandre et al., 2012). By using pairwise distances from the sequence and the prior intra-specific divergence, ABGD analysis detects the gap between intra and interspecific divergence.

The DNA Sequence Polymorphism 6.12.03 software was used to measure nucleotide diversity (p) and haplotype diversity (Hd) in order to analyze genetic diversity in the population of six different locations from Poonch River in a single data set that may help to predict genetic diversity within

**TABLE 2 Comparative analysis of some important phenotypic characters of six fish species habiting River Poonch with already published data.**

| Morphometric/ meristic character | | *S. richardsonii* | *S. plagiostomus* | *T. putitora* | *G. gotyla* | *T. latius* | *G. kashmiriensis* |
|---|---|---|---|---|---|---|---|
| Standard length | Present study | 10.32–31.76 ± 3.2 | 14.2–28.9 ± 3.47 | 17.1–28.8 ± 3.7 | 12.2–18.5 ± 1.50 | 11.1–16.2 ± 1.67 | 6.9–11.2 ± 1.67 |
| | Earlier reported | 19.85 ± 3.25, 15.55–19.06 | 12.4–32.9 ± 3.1 | 16.4–24.2 ± 5.5, 14–32 | 12.1 ± 1.6, 6.6–11.9, 12.1 ± 1.6 | 7.55–13.50 ± 1.93, 7.55–13.50 | 9.2 |
| | References | Khan et al. (2020). Mehmood et al. (2021), Regmi et al. (2021); Lohani et al. (2020) | Mir et al. (2013a); Bhat et al. (2013) | Kamboj & Kamboj, (2019); Mehmood et al. (2021) | Brraich & Akhter, (2015); Attaullah et al. (2021) | Sharma et al. (2014); Mir et al. (2013b) | Wahab and Yousafzai (2017) |
| Maximum body depth | Present study | 3.8–7.5 ± 1.6 | 2.9–6.9 ± 1.8 | 2.9–6.9 ± 1.7 | 1.5–4.9 ± 1.34 | 3.1–4.7 ± 0.43 | 2.6–3.7 ± 0.43 |
| | Earlier reported | 4.95 ± 1.08, 2.69–4.21 | 2.7–5.87 ± 1.18 | 4.1–7.5 ± 1.4, 3.8–8.2 | 1.7–3.6, 2.2 ± 0.1 | 1.91–2.92 ± 0.38 | 2.1 |
| Minimum body depth | Present study | 1.9–3.5 ± 0.7 | 1.6–3.9 ± 1.4 | 2.1–3.8 ± 0.6 | 1.7–2.5 ± 1.04 | 1.3–2.9 ± 0.76 | 1.3–2.3 ± 0.76 |
| | Earlier reported | | 1.45–3.23 | 1.3–4 | 1–1.9 | 0.90–1.70 ± 0.22 | |
| Head length | Present study | 3.3–5.9 ± 0.6 | 2.9–5.7 ± 1.06 | 4.2–7.1 ± 0.9 | 2.9–3.7 ± 0.81 | 2.5–3.2 ± 0.19 | 2.5–3.2 ± 0.19 |
| | Earlier reported | 4.75 ± 0.45 | 2.87–4.98 | 3.5–6.9 | 2–3.8 ± 1.8 | 1.83–2.50 ± 0.22, 1.83–2.50 | 2.5 |
| Dorsal fin rays | Present study | 6–9 | 6–9 | 8–12 | 6–9 | 8–10 | 8–10 |
| | Earlier reported | 8.25 ± 0.48 | 9.38 ± 0.21 | 12 (3/9), 9–13 | 6–10 | III-8 | |
| Pectoral fin rays | Present study | 12–16 | 12–16 | 14–19 | 12–14 | 12–14 | 12–14 |
| | Earlier reported | 18.00 ± 0.91 | 15.31 ± 0.61 | 19, 10–15 | 12–14 (1–6/8–12) | I-13 | |
| Pelvic fin rays | Present study | 9–11 | 9–11 | 8–9 | 8–9 | 8–9 | 8–9 |
| | Earlier reported | 11.50 ± 0.65 | 11.23 ± 0.39 | 9 | 8–9 | I-8 | |
| Anal fin rays | Present study | 5–6 | 5–7 | 5–6 | 6–7 | 5–6 | 5–6 |
| | Earlier reported | 7.75 ± 0.85 | 7.00 ± 0.23 | 2–6, 5–8 | 6 (1–2/4–5) | I-5 | |
| Caudal fin rays | Present study | 16–18 | 16–18 | 16–19 | 16–19 | 16–20 | 16–20 |
| | Earlier reported | 17.9 ± 2.87 | 18.00 ± 0.23 | 19, 15–22 | 17–19 (2–5/13–17) | II-17–18 | |

the population of six endemic fish species found in the region. The first dataset included four species belonging to the family Cyprinidae for the analysis of *Cyt b*, and five species belonging to two families were examined through 16S rRNA in the second subset. While in the third subset, combined studies were performed for both the genes for all the individuals.

# Results

## Phenotypic analysis/identification

The current study successfully identified six endemic freshwater fish species from the Poonch River of the Northwest

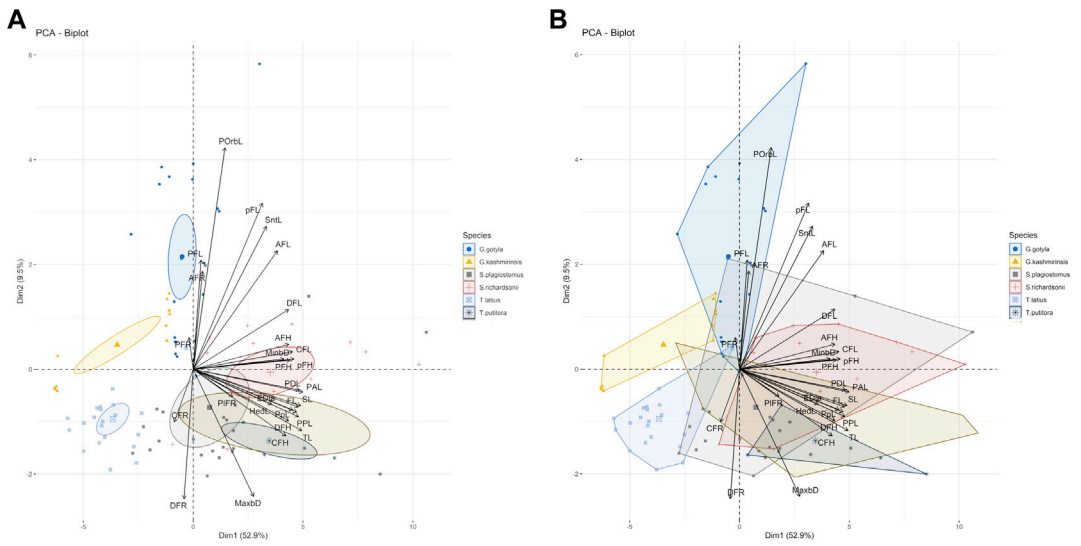



FIGURE 2
**(A,B)** Bi-Plot the variables and species oriented along the first two principal components with species. The Dim 1 and Dim 2 are the first and second principal extracted in PCA and notation in graph refers to the morphometric variables.

Himalayan range of J&K (India), i.e., *S. richardsonii*, *Schizothorax plagiostomus*, *T. putitora*, *G. gotyla*, *T. latius* and *G. kashmirensis*. The species belong to two families and two orders. As per the International Union for Conservation of Nature (IUCN), 2018 status, out of six identified species, *G. kashmirensis* is critically endangered (CR), *T. putitora* endangered (EN), *S. richardsonii* enlisted as vulnerable (VU), *S. plagiostomus* not evaluated (NA), while *G. gotyla* and *T. latius* are listed as least concerned (LC), the same is presented in Tables 1, 2. Multivariate and cluster analyses like ANOVA, PCA, CVA, RDA, dendrogram, and molecular approaches like ABGD, BI, and Phylogenetic Trees were employed to get a robust result. ANOVA demonstrates a significant ($F = 20.90$, $p < .05$) difference in phenotypic measurements and relations in the present study. It is found that by performing the principal component analysis (PCA), 28 principal components were extracted, of which the first two (PC1 and PC2) differentiated 62.4% of the entire variation in terms of variance (52.9 for PC1, and 9.5% for PC2). In terms of influence variables on PC1 the PCA components with the highest loadings are (Total length (TL), Standard length (SL), Forked length (FL), Pre-pectoral length (PPL), Pre-pelvic length (PpL), Pre-dorsal length (PDL), Pre-Anal length (PAL), Pectoral fin height (PFH), pelvic fin height (pFH), Dorsal fin length (DFL), Dorsal fin height (DFH), Anal fin length (AFL), Anal fin height (AFH), Caudal fin length (CFL), Caudal fin height (CFH), Eye diameter (EDia), Head length (HdL), Maximum body depth (MaxbD), Minimum body depth (MinbD). Correspondingly, in PC2, the most crucial loading was Pectoral Fin length (PFL), Pelvic Fin length (pFL), Snout Length (SntL), Pre-Orbital length (PorbL), Dorsal Fin rays (DFR), Anal Fin rays (AFR) and Caudal Fin rays (CFR). The Biplot revealed that all stocks were mixed and overlapped, as shown in Figures 2A, B. According to the PCA plot of PC1 and PC2 (Figure 2B), there is a high overlap between two species of Schizothorax genera. However, morphological variation occurred even in *T. putitora* collected during different season. Species such as *G. kashmirensis* maintained clear differentiation from others. On the whole, PCA separated fish species based on the variation in their outline, regardless of their size, with high overlaps among strictly associated species. As a result of high levels of overlap between groups, further verification was essential through CVA to obtain shape variations. There were 28 canonical variates that accounted for all variances. In the analysis of variance, the first two canonical variables (CV1 and CV2) differentiated 73.68% of the total difference. The importance of CV1 was 45.09% and the importance of CV2 was 28.59%. CVs from two closely related species with overlapping body shapes were observed in the first two CVs (Figures 3A, B). In CVA plots, *T. putitora* collected from different season and *Schizothorax* genera overlapped highly, indicating a high degree of similarity in their shapes and phenotypic characteristics. However, the *G. kashmirinsis* species was placed distinctly from other species groups as this fish belongs to different families. Similar results were obtained through RDA, in which five species were separated at RDA1 and one species (*G. kashmirinsis*) were differentiated at RDA2, as shown in Figure 4A. Moreover, the phenotypic results were conformed through a dendrogram in which all six species were separated successfully shown in Figure 4B.

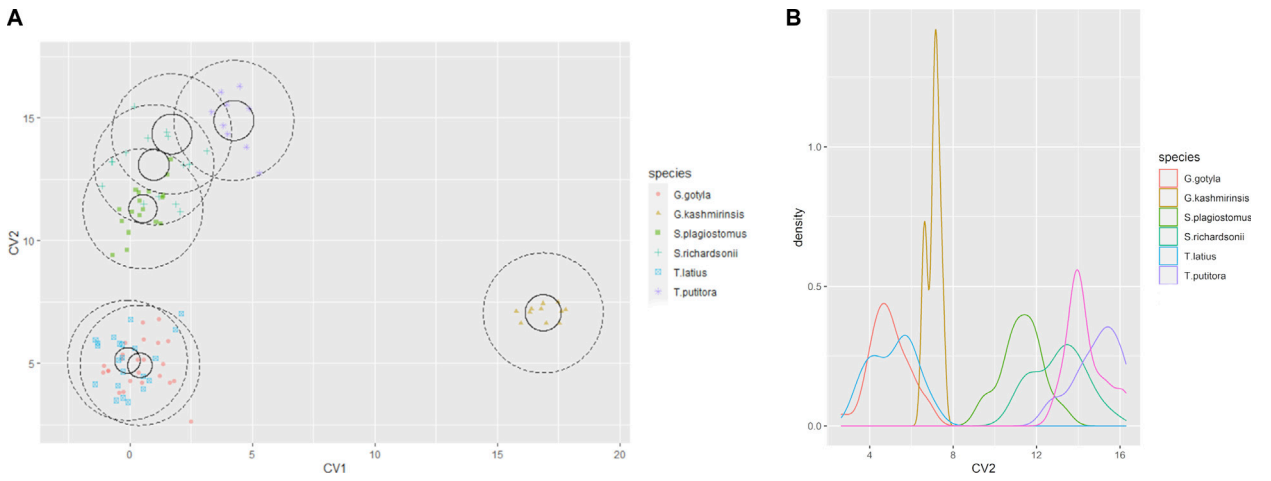



FIGURE 3
**(A)** The Canonical Variate (CVA) showing the morpho-genetic variation of six fish species collected from River Poonch **(B)** biplot of same centroids of the canonical variate score are shown on the 25 morphometric distances.

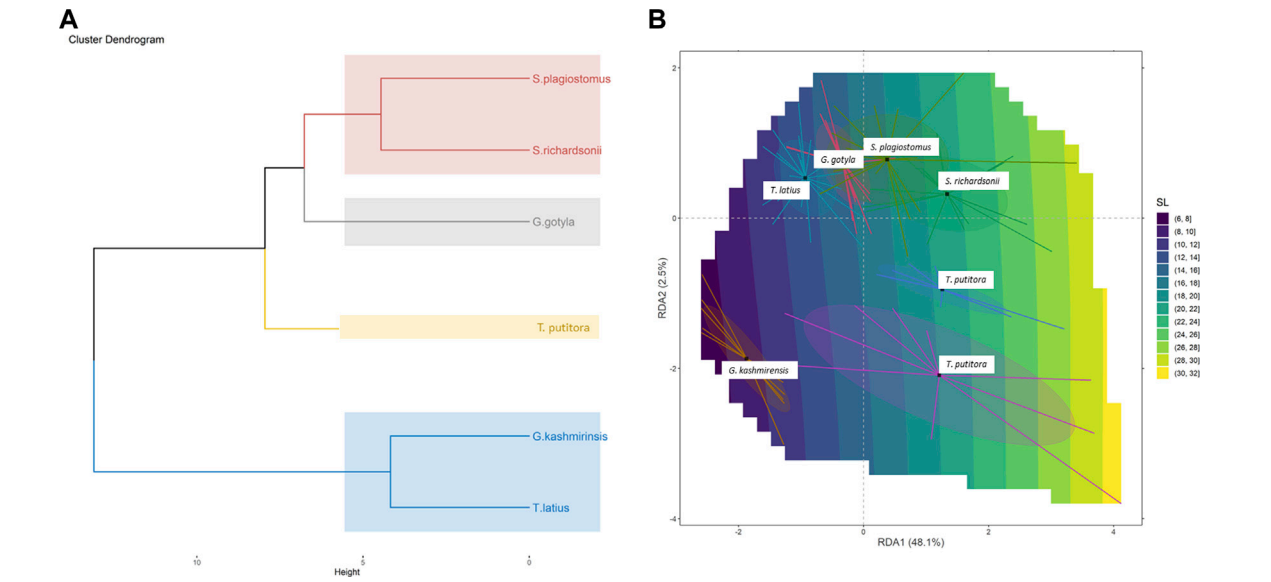



FIGURE 4
**(A)** Redundancy analysis (RDA) Biplot showing taxonomic division among six fish species collected from River Poonch. **(B)** Cluster dendrogram derived from morphometric analysis.

The correlation plot in Figures 5A, B and Supplementary Tables S2, S4–S6 depicted that all phenotypic characters varied uniformly in total length and provided significant and strong positive correlations. The result revealed Head length (HdL), Anal fin length (AFL), Total length (TL), Standard Length (SL), Forked length (FL), Pre-pectoral length (PPL), Pre-pelvic length (PpL), Pre-Dorsal length (PDL), Pre-Anal length (PAL) and Caudal Fin length (CFL) are the most important character that plays a key role in the identification of these fish species in which the value if $p$ is < .001. The highest correlation for *T. putitora* with reverence to total length (Supplementary Tables S1, S3) was observed for forked length ($r$ = .99), pre-anal length ($r$ = .99), standard length

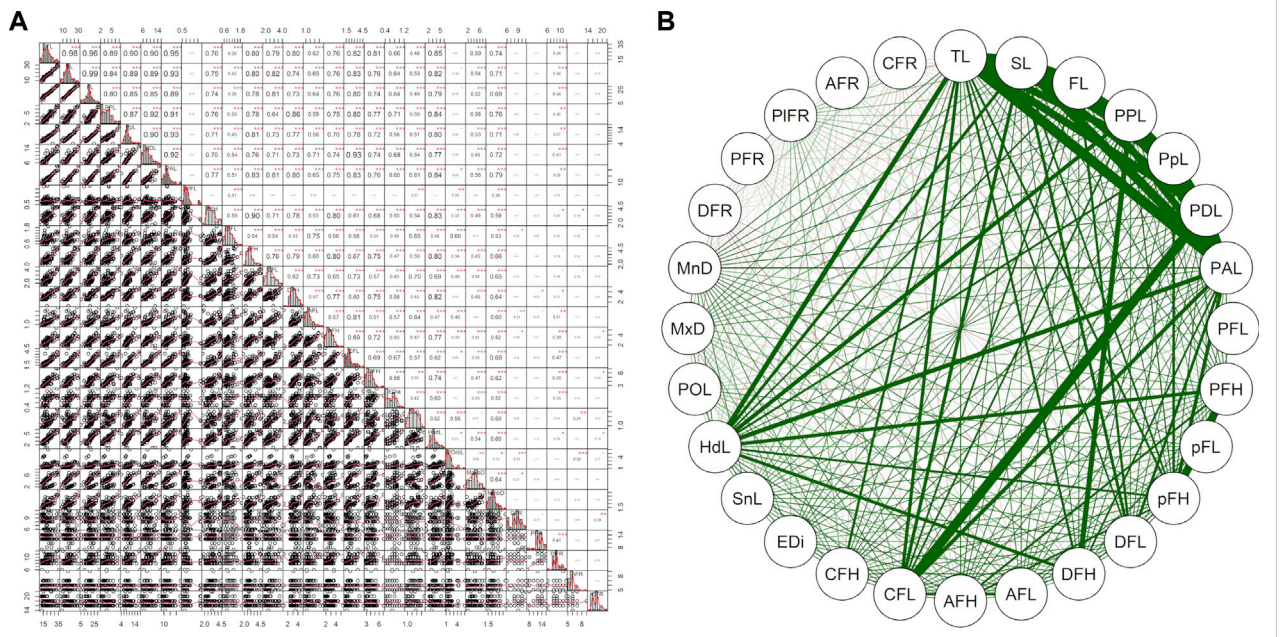



FIGURE 5
**(A)** The correlation plot among the morphometric parameter among six fish species collected from River Poonch. The value given around all the axes are the range of each species parameter s measured unit values (cm). Correlation coefficients (r) are indicated with numeric values, with significant levels (p) are denoted by asterisks (* < .05, ** < .01, *** < .001) and **(B)** presenting same in diagrammatic value.

(*r* = .98), pre-pelvic length (*r* = .98), whereas lowermost was recorded for pelvic fin length (*r* = 52) and caudal fin height (*r* = .55). In the case of head length for *T. putitora*, pre-orbital length (*r* = .77) showed the lowest correlation for head length and the eye diameter (*r* = .52).

Moreover, the present study depicts the pelvic fin length as a diagnostic feature to differentiate *S. richardsonii* from *S. plagiostomus*. However, the highest correlation in overall size was observed for standard length (*r* = .98), forked length (*r* = .98), pre-pelvic fin length (*r* = .96), while the lowest correlation with maximum body depth (*r* = .58), pectoral fin length (*r* = 59). While about head length, the maximum was eye diameter (*r* = .67) and minimum pre-orbital length (*r* = .49) for *S. richardsonii*. The result obtained for *S. plagiostomus* as maximum correlation concerning total length with standard length (*r* = .99), forked size (*r* = .99) pre-pectoral size (*r* = .94), while the least was seen in the case of caudal fin length (*r* = .57), and maximum body depth (*r* = .66). The highest correlation was observed in head length for pre-orbital length (*r* = .52), and the lowest was eye diameter (*r* = .44), as shown in Supplementary Tables S2, S3, S9.

## Molecular analysis

Integrative methods successfully identified all the unpublished *Cyt b* and 16S rRNA sequences from six fish species as it was determined that there was no mismatch between the traditional and advanced techniques used for identifications. The current procedure, morphological and molecular identification, showed a high level of resemblance (97–100%). We obtained mitochondrial barcodes (543 and 519 bp) of *Cyt b* and 16S rRNA, respectively, for six fish species belonging to two families, two orders and five genera collected from six locations in the River Poonch (Figure 1). The amplified sequences did not contain any stop codons, insertions or deletion indicating that all of the segments represent functional mitochondrial sequences. All the fish species except one multiple specimen (minimum of three specimens per species) from different sampling sites were analyzed to document intraspecific divergence. Only one species, i.e., *G. kashmirensis,* were represented by single specimens (Table 1). As specimens were attested to species level based on phylogenetic trees, all morphological identifications improved with molecular identification since the specimens were attributed to species. However, Primers against *Cyt b* used in the present study yield an average nucleotide length of 543 bp, out of which 409 sites were constant, 50 showed singleton variables, 84 were parsimony informative sites and 134 were polymorphic. While in the case of 16S, rRNA 519 bp was obtained. Out of these, 399 sites were constant, 42 were singleton variable sites, 84 were parsimony informative sites, and 95 were polymorphic respectively. All these sequences were submitted to GenBank and got accession numbers presented in

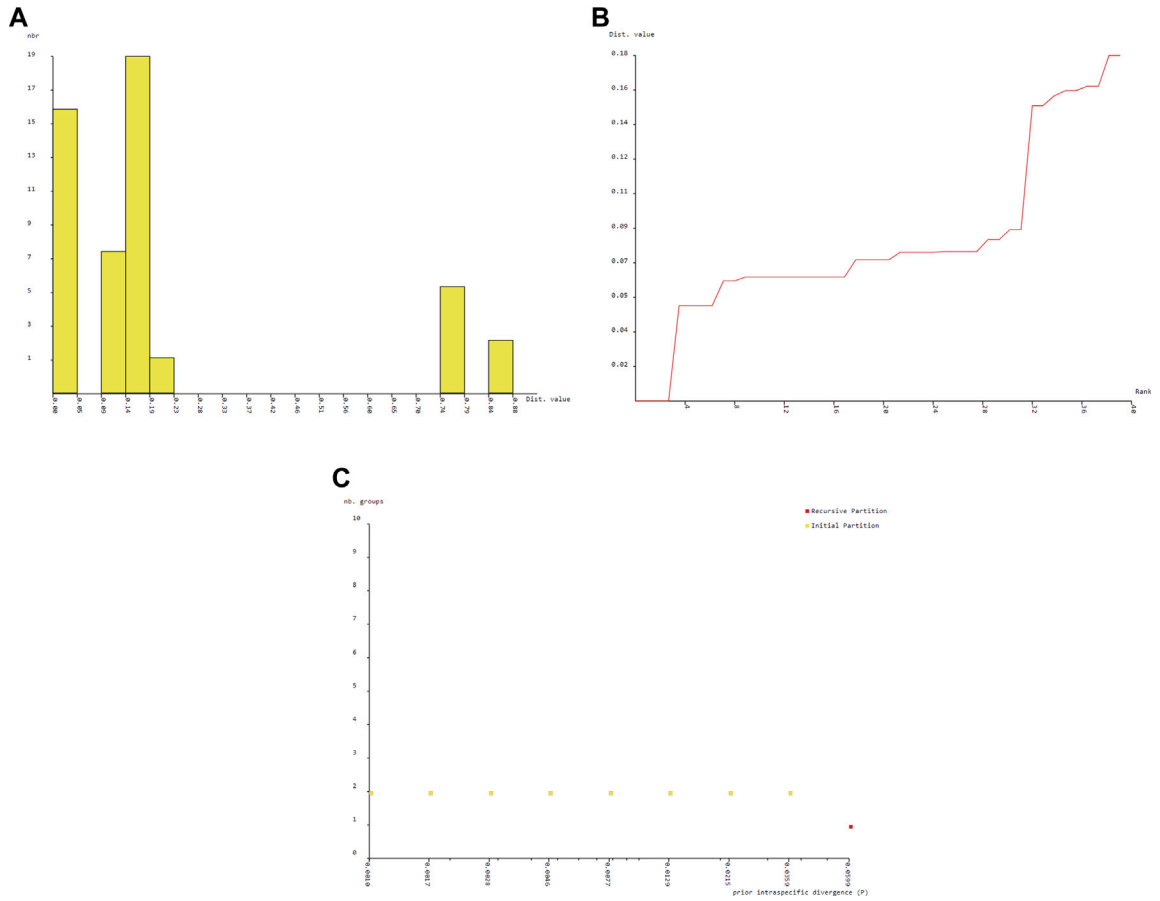



FIGURE 6
Barcode gap analysis of Pangasiid species generated by Automatic Barcode Discovery Gap Discovery (Puillandre et al., 2012). Distributions of K2P distances and between each pair of specimens **(A)** histogram of distance **(B)** ranked distance and **(C)** number of PSHs obtained for each prior intraspecific divergence.

Table 1. Average nucleotide frequencies were A = 28.56%, T/U = 29.75%, C = 26.35% and G = 15.34% for *Cyt b*, while in case of 16S rRNA these were A = 32.39%, T/U = 21.22%, C = 24.51% and G = 21.88%, respectively.

As we expect, in a hierarchy rise in average genetic variant from within species (mean = 014%, standard error [SE] = .000, to within families (mean = .20%, SE = .001) was observed in the K2P model. As a whole, the genomic discrepancy among the same genus was roughly two times more than between the individuals of the same species. It was also noticed that the genetic distances between the lowest to highest taxonomic levels increased as the taxonomic levels increased as DNA sequence analysis of five fish species revealed that the intra-species genetic distance varied from .00 to .14. In contrast, the interspecies genetic distance ranged from .16 to .20. The maximum distance between *T. putitora* and *G. gotyla*, while minimum distances were observed between *T. putitora* and *T. latius*. However, in the case of 16S, the pairwise genetic distance for 16S rRNA sequencing depicted that the intraspecies genetic distance varied from .0 to .6. Similarly, the inter-species genetic distance was also noted from .08 to .16. The analysis also depicted that the maximum distance was observed for *G. kashmirensis* and *T. putitora,* while the minimum was for *S. plagiostomus* and *S. richardsonii* (Supplementary Tables S7, S8).

The BIN analysis led to the recognition of 7 OTUs. All the BIN clusters were found to be taxonomically concordant with the other barcode that was BOLD assigned to the same species name. The count of OTUs produced by ABGD varied from 3 to 7 (Figure 6). The ABGD analysis conducted with K80 Kimura distance with Nb bins (for distance distribution) of 20 and gap widths (X) of .75 produced initial partitions with OUT count of 5 with Barcode gap distance = .069 for *Cyt b* ($p$ = .021544–.012915), 3 in case of 16S rRNA ($p$ = .021544–.012915) and 7 for combined CYT b +16S rRNA ($p$ = .035938–.021544) respectively, whereas the use of P distance returned 5 ($p$ = .021544–.012915), 3 ($p$ =

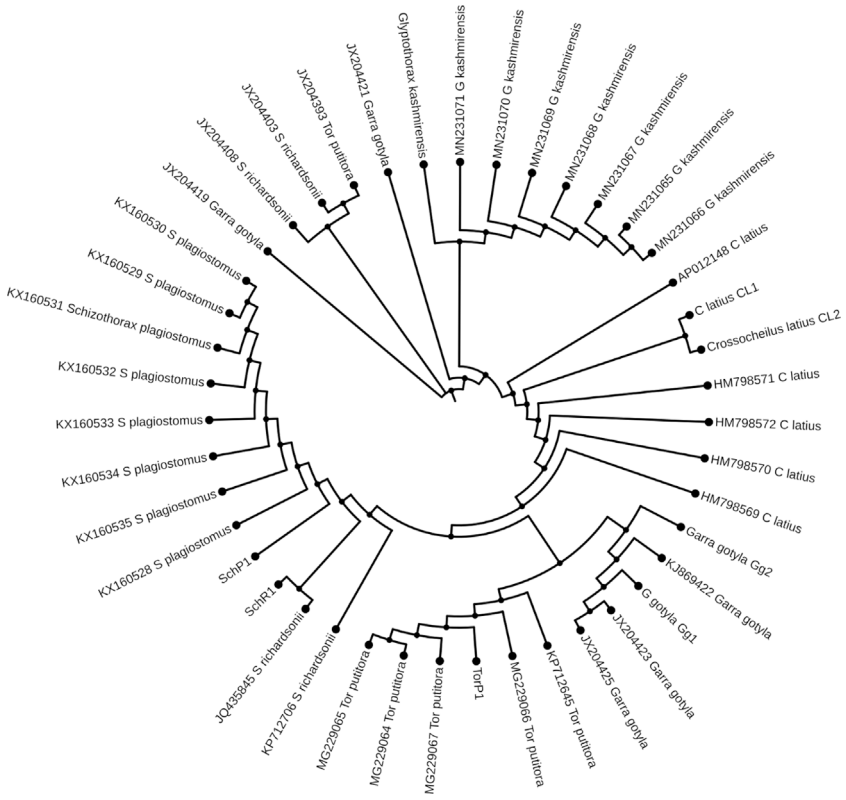



FIGURE 7
Maximum likelihood tree showing relationship of six fish species of River Poonch based on 16S rRNA data. The maximum likelihood tree was generated using GTR+G model.

.021544–.012915) with Barcode gap distance = .493 and 7 ($p$ = .035938–.021544) OTUs with a barcoding gap of .024. A comparison between the Bayesian inference and maximum-likelihood gene tree did not reveal the obvious difference in the positioning of OTUs. The two methods yielded congruent results with slight variation shown in Figure 6.

## Phylogenetic study

The ML and BI trees which were built based on the *Cyt b* and 16S rRNA dataset, placed the reference species grouped in close contact with the species identified by NCBI website, that proves the correctness of our morphological identification (Figures 7, 8). The ML and BI trees grouped sequences of the same taxonomically identified species indicated no overlapping clusters. The topology of ML and BI trees is almost identical. The topologies of *Cyt b* and 16S were slightly different; however, a better result was obtained through a combined approach and therefore, we describe it in detail (Figures 9–11). The result indicated that all six fish species were well differentiated according to their respective position. The tree topology for combined analysis (Figure 11) formed six monophyletic clad representing six species. According to their classification, *G. kashmirensis* form a separate clade, while other species form a clade and sub clade. The genus *Schizothorax* is placed within the same clade next to the *T. putitora*. While in the case of Tor, these are placed under the sister clade immediately next to the *Schizothorax* species.

On the other hand, *G. gotyla* and *T. latius* are placed under a sub-clade closer to the *Schizothorax* genus than *T. putitora* with BS = 81–100; PP = .76). Phylogenetic analyses were also performed separately for *Cyt b* and 16S using PhyML algorithm (parameters—bootstrap value = 100, tree search = "Best from NNI & SPR", Starting tree = "BioNJ," equilibrium and site rate variation = "Optimized"). The better result was obtained through *Cyt b,* in which five monophyletic clades were formed, while on the other hand, only three monophyletic clades were obtained for 16S. Moreover, similar results were obtained through incorporated data as all the species were grouped according to their classification, as shown in Figures 7, 8.

The haplotype networks of the closely related species demonstrated that the sharing of *Cyt b* and 16S rRNA haplotype was common for two genera (*Schizothorax* and

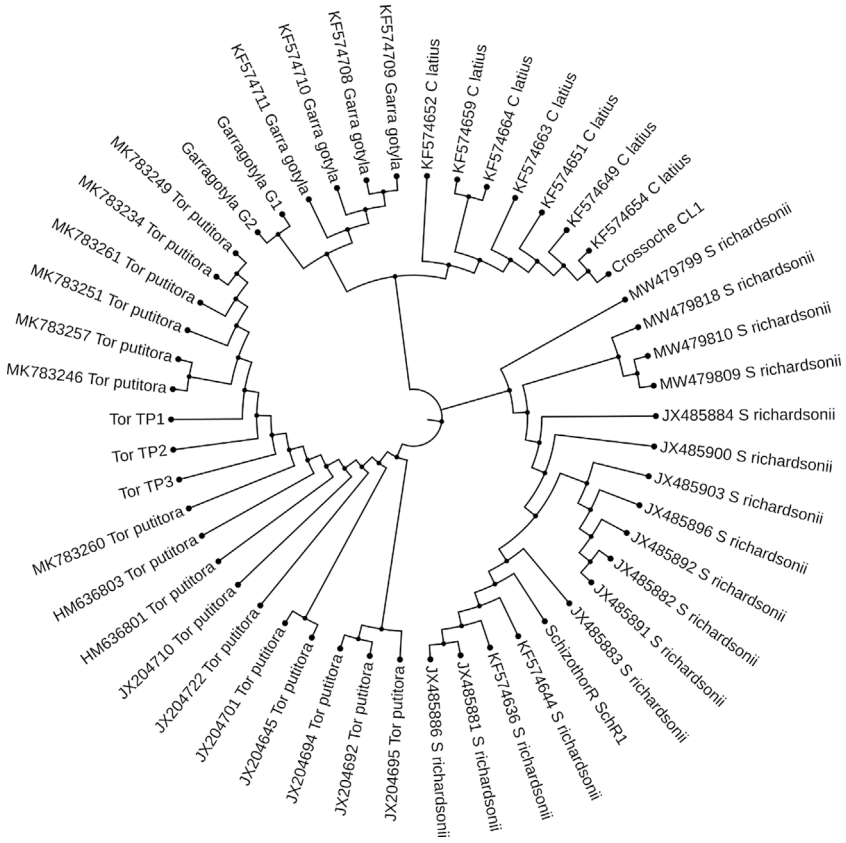



FIGURE 8
Maximum likelihood tree showing relationship of five fish species of River Poonch based on Cyt b data. The maximum likelihood tree was generated using GTR_G model.

*Tor*). The result of the current study noted four haplotypes with .644 haplotype diversity, while nucleotide diversity and overall mean distance were noted as .0904 and .24, respectively, for *Cyt b*. On the other hand, five haplotypes were observed with .899 haplotype diversity, while overall nucleotide diversity and mean found distance were recorded as .0720 and .08. The 16S rRNA sequencing helped us to estimate the average evolutionary divergence and overall sequence pair, which involved ten nucleotide sequences. The DNA sequence information of *Cyt b* gene's able transversion/transition ratio basis (R) was observed as 3.219 and k1 = 4.17 for purine, while k2 = 7.78 for pyrimidines.

# Discussion

## Morphological characteristics

Poonch River is well known for its diverse fish fauna and serves as a lifeline for the large population. Over 25% population of the area is solely dependent on the fishery sector for their livelihood. The current study represents the first-ever effort to combine phenotypic and genotypic analysis of six economically important fish species inhabiting this Himalayan region (Poonch River). The multivariate analysis appears to be a practical approach for identifying and understanding evolutionary relationships among fish species. Therefore, it has gained importance in fishery research (Caillon et al., 2018; Behera et al., 2020). Cyprinid morphological variation, such as body profile, has rarely been studied (Bravi et al., 2013; Yan et al., 2013; Dwivedi et al., 2020). There are still some questions about the fish species inhabiting the Poonch River, and the criteria used to differentiate them have been limited to morphological and meristic characteristics (Day, 1878; Talwar and Jhingran, 1991; Jayaram, 2010). At first glance, after examining specimens morphologically, we were expecting nine fish species during the current study. But after integrative analysis, it was confirmed that only six fish species were there instead of nine. Because *G. gotyla* and *T. latius* were misidentified as *G. lamta* and *T. diplochelius,* respectively, as reported in the previous study by Dutta (2003). Whereas the genus Schizothorax could not be differentiated only on molecular analysis through an individual approach but combined with

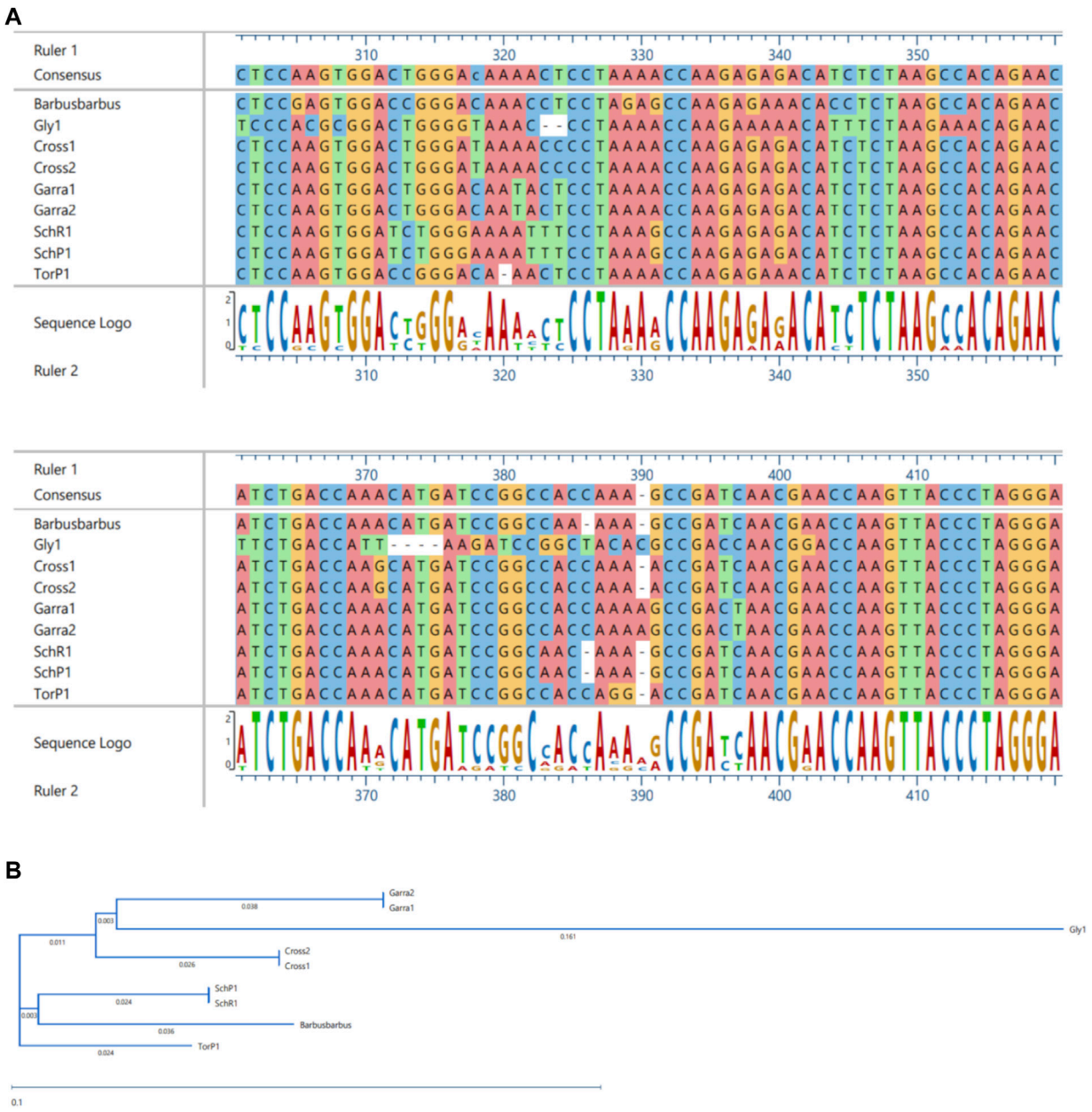



FIGURE 9
MSA **(A)** and Phylogenetic Tree **(B)** of 16S sequences. MSA was performed using MUSCLE Program and Phylogentic analysis were performed using PhyML algorithm (parameters—bootstrap value = 100, tree search = "Best from NNI and SPR," Sarting tree = "BioNJ," equllibrium and site rate variation = "Optimized").

phenotypic study, one can easily distinguish these species (Lal et al., 2015; Dwivedi, 2020). On the other hand, after collection of Tor samples we were expecting two species, *viz.*, *T. putitora* and *Tor tor* as there was huge phenotypic ambiguity in the samples collected from the river, as by employing morphometric analysis it expressed some notable potential differences and on that basis one can easily consider them into two different species of Tor genus. Moreover, some distinguishing features between *T. putitora* and *T. tor* can be seen easily like head length, colour and size of barbells. But molecular results of the current study did not support the phenotypic results as after molecular analysis even not a single nucleotide difference can be sighted. So therefore, we can say that these morphological differences

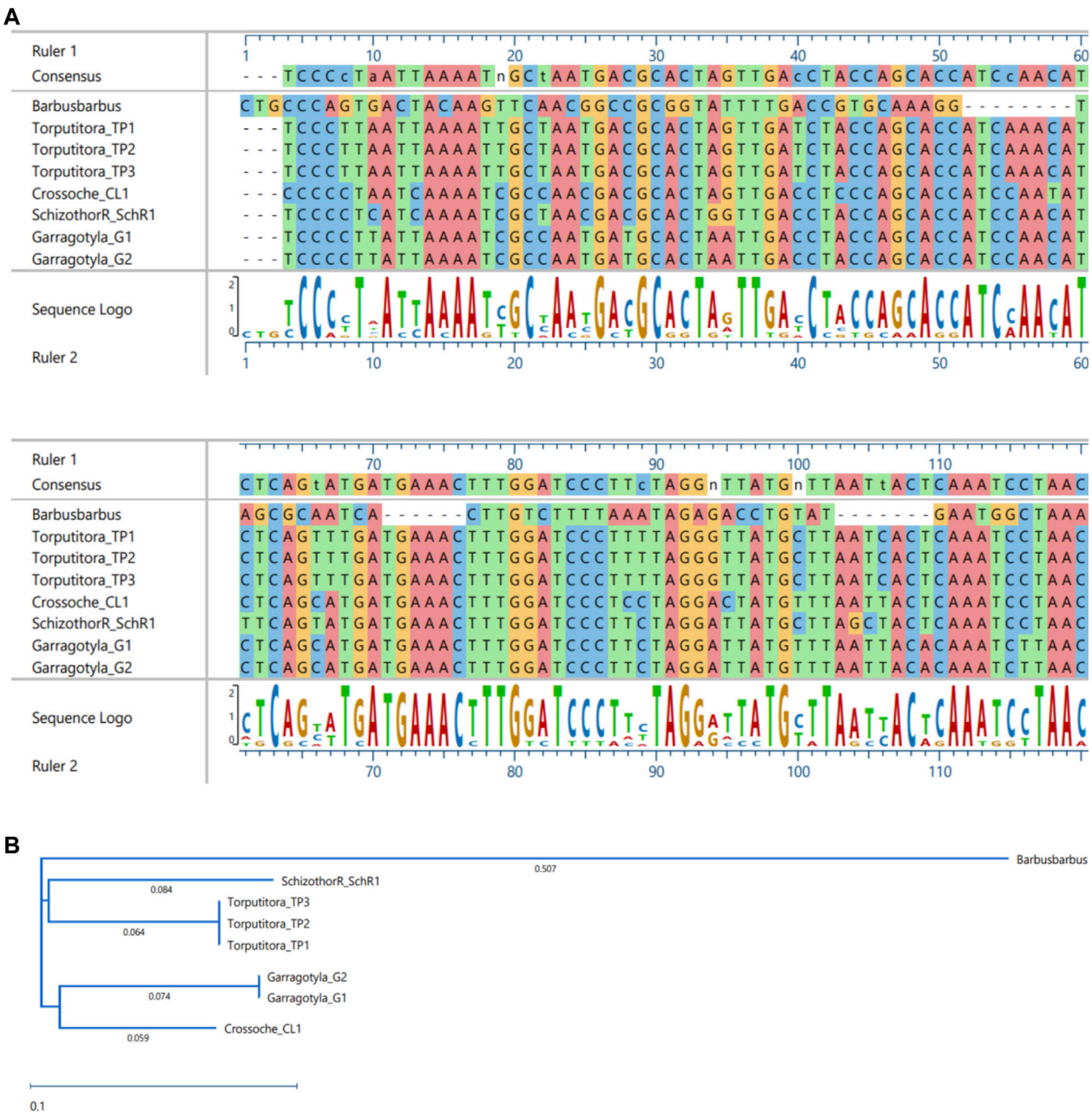



FIGURE 10
MSA **(A)** and Phylogenetic Tree **(B)** of CYT-B sequences. MSA was performed using MUSCLE Program and Phylogentic analysis were performed using PhyML algorithm (parameters—bootstrap value = 100, tree search = "Best from NNI and SPR", Sarting tree = "BioNJ," equllibrium and site rate variation = "Optimized").

are not enough to surpass the molecular result. Hence through integrative approach we concluded that in Tor genus there is only one species i.e., *T. putitora* inhabiting River Poonch. However, such morphometric differences found in samples of Tor genus were to be only due to seasonal difference as the sample collected in different season (Pre-monsoon in which water is crystals clear (March) and during monsoon when the water is turbulent in the month of July as shown in Supplementary Figures S1A, B, respectively). The abovementioned results are in accordance with the findings of Naeem et al. (2011) and Langer et al. (2013), which suggested that

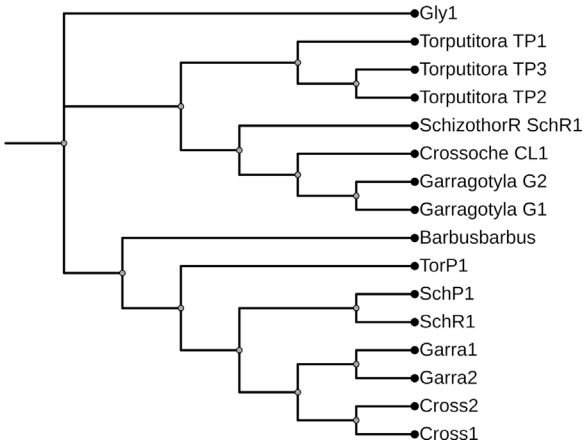



FIGURE 11
Maximum likelihood tree showing relationship of five fish species of River Poonch based on combine data. The maximum likelihood tree was generated using GTR+G model.

*T. putitora* and *T. tor* are morphologically identical and are very difficult to differentiate without having proper taxonomic knowledge.

Moreover, in the current study, *G. lamta* and *T. diplocheilus* could not be differentiated by any approach; hence we conclude that these two species were misidentified from the region. However, Phenotypic analysis employing different approaches like PCA, CVA, RDA, correlation and dendrogram methods (Figures 2–4, 5A, B) among six fish species indicated an unambiguous correlation between the species based on their character. Additionally, three species of the genus Schizothorax and Tor (two and one species respectively) were a cluster in their relevant class, indicating that they share common phenotypic and body patterns and have similar body patterns. On the other hand, *G. kashmirensis* is placed distinctly from other fish species, which indicates that it differs in body profile and is characterized by a special set of features (Figure 3A). The present study's results conform with the findings of other workers (Sarkar et al., 2012; Gupta et al., 2018; Dwivedi et al., 2020).

Based on the results, it was noted that out of six selected fish species, three species, *viz.*, *T. putitora*, *S. richardsonii* and *S. plagiostomus* have similar external morphological characteristics, making them difficult to distinguish. However, a lot of work has been carried out at the taxonomic level of these fish species, but still, there is uncertainty in the identification among these fish species (Pandey and Nautiyal, 1997; Kullander et al., 1999; Dwivedi et al., 2020; Pandey et al., 2020; Jaafar et al., 2021). The result indicated that all the parameters were associated with the total body length and head length. At the same time, variations observed in head length and fin rays became distinctive characteristic features for *T. putitora* (Figures 2, 5A, B). Our findings also revealed that *S. richardsonii* and *S. plagiostomus* are similar in appearance. Similar values were observed for all the parameters except maximum body depth, caudal fin length, and head length. Also, no differences in their habitat and spawning season were observed. The number of inferior pharyngeal teeth in both species was the same (4, 3, 2/2, 3, 4). A lot of work has been reported favoring the current results (Misra et al., 1980; Negi and Negi, 2010; Mir et al., 2013a; Wagle et al., 2015; Wani et al., 2018; Kamboj and Kamboj, 2019). Misra et al. (1980) revealed that both *S. richardsonii* and *S. plagiostomus* had the same number of teeth. Only slight differences in the shape and size of the ceratohyal and epihyal were noted, with the epihyal of *S. richardsonii* being smaller than *S. plagiostomus*. Other similar observations regarding the morphological variations were noted by Wagle et al. (2015), who reported that for *S. richardsonii*, all the parameters were undoubtedly interrelated with total length and head length; the current study also obtained similar results (Figures 5A,B). Thus, it is concluded that the anal fin length and lips resolve the phenotypic problem of differentiating between these two species. Therefore, the anal fin length of *S. plagiostomus* distinguishes it from *S. richardsonii*, as the anal fin of *S. plagiostomus* lying flat and long enough to touch the base of the caudal fin in contrast to the anal fin of *S. richardsonii* which is small, and never touches the base of the caudal fin.

## Molecular analysis (*Cyt b* + 16S rRNA)

Our study marks the first comprehensive molecular evaluation of the six fish species in the Poonch River. In the current study, DNA barcoding was effective in identifying species and provided a straightforward identification system when a perfect match existed between the morphology-based taxonomy and genetic divergence. Overall, this study demonstrated the ability of DNA barcoding to help calibrate the current taxonomic resolution and to shed new light on the fish diversity of River Poonch.

mtDNA is used for phylogenic surveys, genuine identification and differentiation of unidentified or closely linked species of aquatic organisms, including marine and freshwater fishes. A lot of work has been reported in the past on the taxonomy of these fish species from various parts of the world (Dimmick and Edds, 2002; He and Chen, 2006; Thai et al., 2007; Bajpai and Tewari, 2010; Kumar et al., 2011; Lakra et al., 2011; Yang et al., 2012; Ahmad et al., 2014; Chen et al., 2018; Krosch et al., 2019; Ma et al., 2020). However, it was challenging to compare our results with those who studied different species in a single attempt. No work has been reported on these fish species from this water body; however, little work has been reported from other water bodies in the region and other parts of the country (Ahmad et al., 2014; Jaafar et al., 2021), revealing the ambiguity among *Schizothorax* and *Tor* genus. But the current study is different from the previous ones (Ahmad et al., 2014) as they focused on the single genus of *Schizothorax*, but here we

have adopted an integrative approach to five different genera, including *Schizothorax* and *Tor* as well, in a single attempt. Ghouri et al. (2020) reported the efficiency of *Cyt b* for the identification of edible fish species in Pakistan. Through the BLAST tool, they tracked similarities between species. In the current study, we used the ABGD, BINs, ML and BI approach and morphological study to authenticate fish species. However, compared to 16S rRNA, *Cyt b* is more successful in studying the intraspecies phylogenetic relationship. Besides molecular approaches, the morphological study made it more potent for validating the fish species of the area.

Genomics is valuable for preliminary species delimitations and for validating phenotypic-based species circumscription (Puillandre et al., 2012). At first glance, the analysis of *Cyt b* and 16S rRNA sequences with the genetic distance and topology created by the ML and BI tree discriminated all six fish species successfully as no overlapping clusters were formed. However, one species (*G. kashmirensis*) displayed deep divergence and two species of Schizothorax genera displayed the least divergence. A similar result was also found by using the ABGD method, in which the whole data set was delimited into six putative groups with a .24 barcoding gap. However, the individual phylogenetic trees showed overlapping in the case of the *Schizothorax*. *T. putitora, G. gotyla*, *T. latius*, and *G. kashmirensis* could be distinguished by all the methods like ABGD/ML analysis, even with an individual approach. However, the single-gene approach could not distinguish *S. richardsonii, S. plagiostomus*. The phylogenetic tree obtained by 16S rRNA showed a close relationship between the genus *Schizothorax* and formed only three monophyletic clades with three putative groups with three bins. On the other hand, in the case of *Cyt b,* there were five monophyletic clades formed with five putative groups in which five bins were present. However, in the case of *Cyt b*, better results were obtained as species of the Tor genus were partially differentiated. But in combined data that include both *Cyt b* and 16S in a single data set, all the species, including the genus of *Schizothorax*, could be distinguished. The current result was partially in accordance with the previous result of Ahmad et al. (2014), in which they emphasized more on combined analysis and got better results. However, in the case of our result, *G. gotyla* and *T. latius* shared a common sister clade and are also parallel with the result of Pandey et al. (2020), in which they found a similar type of result. Moreover, the current result indicated that the integrative approach could only be an effective tool to remove the ambiguity in these species.

During the current study, *G. gotyla*, *T. latius* and *G. kashmirensis* were clustered symmetrically in their respective genera. However, *G. gotyla* and *T. latius* were grouped and formed a separate sub-clade in all the cases, as both species belong to the sub-family Garrianae. Contrary to this, *G. kashmirensis* formed a separate clade both in individual (*Cyt b*) and (16S) as well as combined (*Cyt b* + 16S) phylogenetic analysis because of the difference in the family as it belongs to the family Sisordae. Conversely, all other species belong to the same family (Cyprinidae) and thus create a monophyletic clade according to their evolutionary history. But genus *Schizothorax* was not adequately differentiated with *Cyt b* as well as 16S individually. The current result was comparable with those of some earlier workers, as they have also noted the minimum genetic difference between these two species (Laskar et al., 2013; Khare et al., 2014; Yang et al., 2015; Laskar et al., 2018). In the present study, the phylogenetic results revealed a close relationship between *S. richardsonii* and *S. plagiostomus,* which supports the findings of other workers (Khan et al., 2016; Ma et al., 2020; Rehman et al., 2020). On the other hand, it has been noticed that there is a close association between the genus *Schizothorax* and *Tor* than *Garra* and *Tariqilabeo*. In general, similar species have been clustered together and different species have formed their cladogram (Lakra et al., 2011; Basheer et al., 2015; Bineesh et al., 2015; Davassy et al., 2015; Rathipriya et al., 2019). However, the considerable variability in their respective sites might be due to the *G. kashmirensis*, which belongs to Sisordae, whereas other species belong to the same family. Moreover, we found overall haplotype diversity higher in 16S rRNA compared to the *Cyt b* gene. In contrast, nucleotide diversity was noted more in the case of *Cyt b* than 16S rRNA. More haplotype diversity indicated that the diversity of this particular river is very high. Dekui et al. (2004) noted a similar result, who pointed out that the *Cyt b* gene revealed 659 constant sites, 481 variable sites, and 419 informative sites. The present results also showed similarity with the study of Lv et al. (2018), who obtained 42.7% of AT content during their analysis for fish species belonging to the family Sisordae.

For both genes, the analysis of the transition transversion ratio for all the species showed a higher transition ratio as compared to transversion, which is in agreement with the study of different authors (Page and Holmes, 1998; Ward et al., 2005; Li et al., 2013; Chakraborty et al., 2014; Bashir et al., 2016; Sharma et al., 2016). Our results also revealed that the average mean distance for *Cyt b* was noted in the range from .00 to .14 for intra-species, whereas .16 to .20 for interspecies with .01 standard error. The maximum distance was observed between *T. putitora* and *G. gotyla*, while the minimum was between *S. richardsonii* and *S. plagiostomus*. Comparable results were obtained from the 16S rRNA sequencing study, which revealed that within the species, variations varied from .00 to .6, but between species, variations ranged from .08 to .16. Moreover, the mean distance of individual species was noted as .00 to .041674 for *G. gotyla,* .00 to .026676 for *T. latius,* .00 to .018240 in the case of *S. plagiostomus* while for *S. richardsonii* it was varied from .00 to .027842, while, for *T. putitora* range from .00 to .022429 and for *G. kashmirensis* it was .075259. However, the extreme distance between *G. kashmirensis* and *T. putitora* was noted, while the least distance was observed for *S. plagiostomus* and *S. richardsonii*. The present study highlighted that intraspecies nucleotide diversity is far lower

than inter-species genetic diversity, according to Ward et al. (2005) and Lakra et al. (2011). At the same time, Sherzada et al. (2020) also noted a gradual increase in genetic diversity with the rise in the genetic level. In the current study, the cases of weak genetic differentiation or haplotype sharing include four species from two genera. The accuracy of DNA barcoding to separate the species depends on the level of genetic diversity (Meyer and Paulay, 2005). Therefore, the incompetence to separate these species is credited to absence of sufficient genetic discrepancy between these species. In the genus *Schizothorax*, interspecific haplotype sharing is an omnipresent example in the same drainage (Dimmick and Edds, 2002; He and Chen, 2006). First, species inhabiting the same waterbody with a large allocation range can display structural variations like the shape of the mouth and lips in case of *Schizothorax*, head length, and body depth for *T. putitora* characteristics for identification (Wu and Wu, 1992; Chen et al., 2015).

## Conclusion

The integrative approach, i.e., morphometric and molecular phylogeny, is the most authentic and informative approach used to discriminate fishes (Rajpoot et al., 2016). In the current study, two mitochondrial genes were used to validate the fish species using a binary approach, such as phenotypic and genotypic data in order to resolve the ambiguity among six important fish species of the Poonch River. A combined analysis of mitochondrial loci mostly agrees with morpho taxonomical studies (Silas, 1960; Ahmad et al., 2014), indicating that the integrative method could effectively resolve uncertainty in the identification and understanding of morphological relationships even with a low sample size. Therefore, molecular phylogenies and phenotypic studies can be compared using standard methods to come up with a robust conclusion. The study showed that *Cyt b* and 16S rRNA were effective indicators of interspecies relationships and genetic variation. Although, we found that all other species, including *G. kashmirensis* (critically endangered), *T. putitora* (endangered), and *S. richardsonii* (vulnerable), are relatively large populations. However, they are still under intense pressure from human activities such as illegal fishing and killing species regardless of their conservation value.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, MW191577; https://www.ncbi.nlm.nih.gov/genbank/, MW191578; https://www.ncbi.nlm.nih.gov/genbank/, MW191579; https://www.ncbi.nlm.nih.gov/genbank/, MW191580; https://www.ncbi.nlm.nih.gov/genbank/, MW191581; https://www.ncbi.nlm.nih.gov/genbank/, MW191586; https://www.ncbi.nlm.nih.gov/genbank/, MW191585; https://www.ncbi.nlm.nih.gov/genbank/, MW148582; https://www.ncbi.nlm.nih.gov/genbank/, MW148583; https://www.ncbi.nlm.nih.gov/genbank/, MW148584; https://www.ncbi.nlm.nih.gov/genbank/, MW148588; https://www.ncbi.nlm.nih.gov/genbank/, MW148587; https://www.ncbi.nlm.nih.gov/genbank/, MW148589; https://www.ncbi.nlm.nih.gov/genbank/, MW148585; https://www.ncbi.nlm.nih.gov/genbank/, MW148586.

## Ethics statement

The animal study was reviewed and approved by a full compliance with the relevant international, national, and institutional guidelines for the use and care of animals was observed during the current research project. Animal Ethical Committee has approved all the protocols used registered under R. No. 801/Go/RE/S/2003/CPCSEA.

## Author contributions

MA performed the experimental work and wrote the manuscript. IA conceived, supervised and revised the manuscript. SA designed the molecular work and helped in statistical analysis. KA-A help in drafting of the manuscript and corrections, MF also help in drafting of the manuscript and corrections. BB helped in statistical analysis. All authors read and approved the final manuscript.

## Funding

This work was supported by the Researchers Supporting Project number (RSP2023R154), King Saud University, Riyadh, Saudi Arabia.

## Acknowledgments

The authors would like to express their sincere gratitude to the Head, Department of Zoology, University of Kashmir, Hazratbal, Srinagar, India, who generously provided the laboratory facilities. Also thankful to the Head, Division of Animal Biotechnology, Faculty of Veterinary Sciences and Animal Husbandry, Shuhama, SKUAST-K for carrying out the manuscript's molecular part. The authors would like to extend their sincere appreciation to the Researchers Supporting Project number (RSP2023R154), King Saud University, Riyadh, Saudi Arabia.

## Conflict of interest

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1047436/full#supplementary-material

**SUPPLEMENTARY FIGURE S1**
Representative images of each species **(A)** *T. putitora* **(B)** *T. tor* **(C)** *G. gotyla* **(D)** *S. richardsonii* **(E)** *S. plagiostomus* **(F)** *T. latius* **(G)** *G. kashmirensis*.

**SUPPLEMENTARY FIGURE S2**
Location of 14 landmarks refers to 1. Anterior tip of snout at upper jaw 2 most posterior espect of neurocranium 3. Origin of dorsal; fin 4. Anterior attachment of dorsal membrane from caudal fin 5. Posterior end of vertebral column 6. Anterior attachment of ventral membrane from caudal fin 7. Origin of anal fin 8. Insertion of pelvic fin 9. Ventral edge of the operculum 10. Origen of pectoral fin 11. Posterior end of opperculum 12. Anterior end of eye 13. Posterior end of eye 14. Fork point of caudal fin.

## References

Ahmad, S. M., Bhat, F. A., Balkhi, M. U. H., and Bhat, B. A. (2014). Mitochondrial DNA variability to explore the relationship complexity of *Schizothoracine* (Teleostei: Cyprinidae). *Genetica* 14, 507–516. doi:10.1007/s10709-014-9797-y

Ahmed, I., Ahmad, I., Dar, S. A., Awas, M., Kaur, H., Ganai, B. A., et al. (2019). *Myxobolus himalayaensis* sp. nov.(Cnidaria: Myxozoa) parasiting *Schizothorax richardsonii* (cyprinidae: Schizothoracinae) from River Poonch in north west Himalaya, India. *Aquac. Rep.* 14, 100192. doi:10.1016/j.aqrep.2019.100192

Attaullah, U., Ullah, U., Ilahi, I., Ahmad, N., Rahman, F. U., Ullah, J., et al. (2021). Taxonomic, morphometric and limnological assessment of the commercially important ichthyofauna of Sakhakot Stream, Malakand, Pakistan. *Braz J Biol.* 82, e243774. doi:10.1590/1519-6984.243774

Awas, M., Ahmed, I., and Sheikh, Z. A. (2020). Length-weight relationship of six coldwater food fish species of River Poonch, Pir Panjal Himalaya, India. *Egypt. J. Aquat. Biol. Fish.* 24 (2), 353–359. doi:10.21608/ejabf.2020.82230

Badkur, R., and Parashar, A. (2015). Morphometric approach towards growth performance of Mahseer (Tor tor) in river Narmada near Hoshangabad (MP). *Indian J. Pharm. Biol. Res.* 3 (2), 66–72. doi:10.30750/ijpbr.3.2.7

Bajpai, N., and Tewari, R. R. (2010). Mitochondrial DNA sequence-based phylogenetic relationship among flesh flies of the genus *sarocophaga* (sarcophagidae: Diptera). *J. Genet.* 89 (1), 51–54. doi:10.1007/s12041-010-0010-5

Basheer, V. S., Mohitha, C., Vineesh, N., Divya, P. R., Gopalakrishnan, A., and Jena, J. K. (2015). Molecular phylogenetics of three species of the genus Rastrelliger using mitochondrial DNA markers. *Mol. Biol. Rep.* 42, 873–879. doi:10.1007/s11033-014-3710-8

Bashir, A., Bisht, B. S., Mir, J. I., Patiyal, R. S., and Kumar, R. (2016). Morphometric variation and molecular characterization of snow trout species from Kashmir valley, India. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 27, 4492–4497. doi:10.3109/19401736.2015.1101537

Bavinck, M., and Johnson, D. (2008). Handling the legacy of the blue revolution in India-social justice and small-scale fisheries in a negative growth scenario. *Am. Fish. Soc. Symp.* 49 (1), 585–599.

Behera, P. R., Jishnudev, M. A., Saravanan, R., Roul, S. K., Ghosh, S., Mahesh, V. U., et al. (2020). Redescription of the enigmatic jellyfish, Crambionella annandalei (Cnidaria: Scyphozoa) from Indian waters. *J. Mar. Biol. Assoc. U. K.* 100, 691–699. doi:10.1017/S0025315420000703

Bhat, F. A., Balkhi, M. H., Najar, A. M., and Yousuf, A. R. (2013). Distribution pattern, density and morphometric characteristics of Shizothoracines (Snow trout) in Lidder River, Kashmir. *Biosean* 8 (2), 363–379.

Bineesh, K. K., Mohitha, C., Vineesh, N., Basheer, V. S., Joselet, M., Pillai, N. K., et al. (2015). Molecular identification of three deepsea fish species of the genus Chelidoperca (Perciformes: Serranidae) from Indian waters. *Indian J. Fish.* 62, 104–108.

Bottero, M. T., and Dalmasso, A. (2011). Animal species identification in food products: Evolution of biomolecular methods. *Vet. J.* 190, 34–38. doi:10.1016/j.tvjl.2010.09.024

Bravi, R., Ruffini, M., and Scalici, M. (2013). Morphological variation in riverine cyprinids: A geometric morphometric contribution. *Ital. J. Zool. (Modena).* 80, 536–546. doi:10.1080/11250003.2013.829129

Briolay, J., Galtier, N, Brito, R. M., and Bouvet, Y. (1998). Molecular phylogeny of Cyprinidae inferred from cytochrome b DNA sequences. *Mol. Phylogenet. Evol.* 9, 100–108. doi:10.1006/mpev.1997.0441

Brown, C., Zakaria, V., Joubert, A., Rafique, M., Murad, J., King, J., et al. (2019). Achieving an environmentally sustainable outcome for the gulpur hydropower project in the Poonch River mahaseer national park, Pakistan. *Sustain. Water Resour. Manag.* 5 (2), 611–628. doi:10.1007/s40899-018-0227-7

Brraich, O. S., and Akhter, S. (2015). Morphometric characters and meristic counts of a fish, *Garra gotyla gotyla* (Gray) from Ranjit Sagar Wetland, situated in the Himalayan foothills, India. *Int. Res. J. Biol. Sci.* 4 (5), 66–72.

Caillon, F., Bonhomme, V., Mvollmann, C., and Frelat, R. (2018). A morphometric dive into fish diversity. *Ecosphere* 9, e02220. doi:10.1002/ecs2.2220

Chakraborty, M., and Ghosh, S. K. (2014). An assessment of the DNA barcodes of Indian freshwater fishes. *Gene* 537, 20–28. doi:10.1016/j.gene.2013.12.047

Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *J. Stat. Softw.* 61 (6), 1–36. doi:10.18637/jss.v061.i06

Chen, W., Ma, X., Shen, Y., Mao, Y., and He, S. (2015). The fish diversity in the upper reaches of the Salween River, Nujiang River, revealed by DNA barcoding. *Sci. Rep.* 5 (1), 17437. doi:10.1038/srep17437

Chen, Y. Y., Li, R., Li, C. Q., Li, W. X., Yang, H. F., Xiao, H., et al. (2018). Testing the validity of two putative sympatric species from Sinocyclocheilus (Cypriniformes: Cyprinidae) based on mitochondrial cytochrome b sequences. *Zootaxa* 4476, 130–140. doi:10.11646/zootaxa.4476.1.12

Cornetti, L., Fields, P. D., Van Damme, K., and Ebert, D. (2019). A fossil-calibrated phylogenomic analysis of Daphnia and the Daphniidae. *Mol. Phylogenet. Evol.* 137, 250–262. doi:10.1016/j.ympev.2019.05.018

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* 9 (8), 772. doi:10.1038/nmeth.2109

Davassy, A., Kumar, R., Shajitha, P. P., John, R., Padmakumar, K. G., Basheer, V. S., et al. (2015). Genetic identification and phylogenetic relationships of Indian clariids based on mitochondrial COI sequences. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 27 (5), 3777–3780. doi:10.3109/19401736.2015.1079901

Day, F. (1878). *The fishes of India being a natural history of the fishes known to inhabit the seas and fresh waters of India, Burma and ceylon, vol. 2.* New Delhi, India: Today & Tomorrow's book agency.

Dekui, H. E., Chen, Y., Chen, Y., and Chen, Z. (2004). Molecular phylogeny of the specialized Schizothoracine fishes (Teleostei: Cyprinidae), with their implications for the uplift of the Qinghai- Tibetan Plateau. *Chin. Sci. Bull.* 49 (1), 39–48. doi:10.1360/03wc0212

Diaz, J., Villanova, G. V., Brancolini, F., Del Pazo, F., Posner, V. M., Grimberg, A., et al. (2016). First DNA barcode reference library for the identification of South American freshwater fish from the lower Parana River. *PLoS One* 11, e0157419. doi:10.1371/journal.pone.0157419

Dimmick, W. W., and Edds, D. R. (2002). Evolutionary genetics of the endemic Schizorathicine Cypriniformes: Cyprinidae) fishes of lake rara, Nepal. *Biochem. Syst. Ecol.* 30, 919–929. doi:10.1016/s0305-1978(02)00030-3

Dutta, S. P. S. (2003). Fish fauna of Poonch district, Jammu region (J&K state). *J. Aquacult. Biol.* 4 (2), 241–246.

Dwivedi, A. K. (2020). Differentiating three Indian shads by applying shape analysis from digital images. *J. Fish. Biol.* 96, 1298–1308. doi:10.1111/jfb.14074

Dwivedi, A. K., Verma, H., Dewan, S., and Verma, S. (2020). Trends in morphological relationships among Gangetic Cyprinids. *Zool. Anz.* 289, 123–132. doi:10.1016/j.jcz.2020.10.003

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32 (5), 1792–1797. doi:10.1093/nar/gkh340

FAO (2000). *The state of world fisheries and aquaculture 2000*. Rome, Italy: FAO.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17 (6), 368–376. doi:10.1007/BF01734359

Froese, R., and Pauly, D. (2007). Fishbase: World wide web electronic publication. Available at: http://www.fishbase.org.

Ghosh, A. K. (1997). "Himalayan fauna with special reference to endangered and endemic species," in *Himalayan biodiversity: Action plan*. Editor U. Dhar (Kosi-Katarmal, Almora: GB Pant Institute of Himalayan Environment & Development), 53–59.

Ghouri, M. Z., Ismail, M., Javed, M. A., Khan, S. H., Munawar, N., Umar, A. B., et al. (2020). Identification of edible fish species of Pakistan through DNA barcoding. *Front. Mar. Sci.* 868. doi:10.3389/fmars.2020.554183

Gonzales, E., and Sartori, J. S. (2002). *Crescimento e Metabolismo Muscular. Fisiologia Aviária Aplicada a Frangos de Corte*. Jaboticabal, SP, Brazil: FUNEP/UNESP, 279-29.

Gonzalez, V. H., Griswold, T., and Engel, M. S. (2013). Obtaining a better taxonomic understanding of native bees: Where do we start? *Syst. Entomol.* 38, 645–653. doi:10.1111/syen.12029

Gopi, K., and Mishra, S. (2014). "Diversity of marine fish of India," in *Marine faunal diversity in India*, 171–193.

Griffiths, A. M., Miller, D. D., Egan, A., Fox, J., Greenfield, A., and Mariani, S. (2013). DNA barcoding unveils skate (Chondrichthyes: Rajidae) species diversity in 'ray' products sold across Ireland and the UK. *PeerJ* 1, e129. doi:10.7717/peerj.129

Guindon, S., Delsuc, F., Dufayard, J. F., and Gascuel, O. (2009). "Estimating maximum likelihood phylogenies with PhyML," in *Bioinformatics for DNA sequence analysis* (Totowa, NJ, USA: Humana Press), 113–137.

Gupta, D., Dwivedi, A. K., and Tripathi, M. (2018). Taxonomic validation of five fish species of subfamily Barbinae from the Ganga river system of northern India using traditional and truss analyses. *PloS One* 13, e0206031. doi:10.1371/journal.pone.0206031

He, D., and Chen, Y. (2006). Biogeography and molecular phylogeny of the genus Schizothorax (Teleostei: Cyprinidae) in China inferred from cytochrome b sequences. *J. Biogeogr.* 33, 1448–1460. doi:10.1111/j.1365-2699.2006.01510.x

Hebert, P. D., Cywinska, A., Ball, S. L., and deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270, 313–321. doi:10.1098/rspb.2002.2218

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17 (8), 754–755. doi:10.1093/bioinformatics/17.8.754

Hussain, S., and Dutta, S. P. S. (2016). Fish fauna of tehsil mendhar, district-poonch, Jammu region, J&K state, with a new record of *botiabirdi, ompokpabda* and *Glyptothorax punjabiensis* from Poonch district. *Proc. Zool. Soc. Ind.* 15, 1–11.

International Union for Conservation of Nature (IUCN) (2018). *Red list of threatened species IUCN*.

Jaafar, F., Na-Nakorn, U., Srisapoome, P., Amornsakun, T., Duong, T. Y., Gonzales-Plasus, M. M., et al. (2021). A current update on the distribution, morphological features, and genetic identity of the southeast asian mahseers, tor species. *Biology* 10 (4), 286. doi:10.3390/biology10040286

Jayaram, K. C. (2010). *The freshwater fishes of the Indian region*. New Delhi: Narendra Publication Company., 616.

Johnson, R. A., and Wichern, D. W. (1998). *Applied multivariate statistical analysis*. 4 ed. New York, USA: Prentice-Hall.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika* 39, 31–36. doi:10.1007/bf02291575

Kamboj, N., and Kamboj, V. (2019). Morphometric and meristic study of four freshwater fish species of river Ganga. *Indian J. Animal Sci.* 89 (4), 120–123.

Kartavtsev, Y. P., Batischeva, N. M., Bogutskaya, N. G., Katugina, A. O., and Hanzawa, N. (2017). Molecular systematics and DNA barcoding of Altai Osmans, oreoleuciscus (pisces, cyprinidae, and leuciscinae), and their nearest relatives, inferred from sequences of cytochrome b (Cyt-b), cytochrome oxidase c (Co-1), and complete mitochondrial genome. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 28 (4), 502–517. doi:10.3109/24701394.2016.1149822

Kassambara, A., and Mundt, F. (2020). *factoextra: Extract and visualize the results of multivariate data analyses*. R package version 1.0.7. Available at: https://CRAN.R-project.org/package=factoextra.

Kassambara, A. (2017). *Practical guide to principal component methods in R: PCA, M(CA), FAMD, MFA, HCPC, factoextra (STHDA)*.

Keskin, E., and Atar, H. H. (2013). DNA barcoding commercially important fish species of Turkey. *Mol. Ecol. Resour.* 13, 788–797. doi:10.1111/1755-0998.12120

Khan, M., Nasir Khan Khattak, M., He, D., Liang, Y., Li, C., UllahDawar, F., et al. (2016). The complete mitochondrial genome organization of *Schizothorax plagiostomus* (Teleostei: Cyprinidae) from Northern Pakistan. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 27 (5), 3630–3632. doi:10.3109/19401736.2015.1079829

Khan, M. A., Khan, S., and Miyan, K. (2012). Length–weight relationship of giant snakehead, Channa marulius and stinging catfish, Heteropneustes fossilis from the River Ganga, India. *J. Appl. Ichthyol.* 28 (1), 154–155. doi:10.1111/j.1439-0426.2011.01901.x

Khan, M. R., Rakha, B. A., and Ansari, M. S. (2020). Diversity and distribution of genera Schizothorax and schizothorichthyes in river swat, Pakistan. *Pak. J. Zool.* 52 (6), 2419. doi:10.17582/journal.pjz/20180904060941

Khare, P., Mohindra, V., Barman, A. S., Singh, R. K., and Lal, K. K. (2014). Molecular evidence to reconcile taxonomic instability in mahseer species (Pisces: Cyprinidae) of India. *Org. Divers. Evol.* 14 (3), 307–326. doi:10.1007/s13127-014-0172-8

Kimura, M. A. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16 (2), 111–120. doi:10.1007/BF01731581

Knebelsberger, T., Landi, M., Neumann, H., Kloppmann, M., Sell, A. F., Campbell, P. D., et al. (2014). A reliable DNA barcode reference library for the identification of the north European shelf fish fauna. *Mol. Ecol. Resour.* 14, 1060–1071. doi:10.1111/1755-0998.12238

Kochzius, M., Seidel, C., Antoniou, A., Botla, S. K., Campo, D., Cariani, A., et al. (2010). Identifying fishes through DNA barcodes and microarrays. *PLoS One* 5, e12620. doi:10.1371/journal.pone.0012620

Kocovsky, P. M., Adams, J. V., and Bronte, C. R. (2009). The effect of sample size on the stability of principal components analysis of truss-based fish morphometrics. *Trans. Am. Fish. Soc.* 138, 487–496. doi:10.1577/t08-091.1

Krosch, M. N., Schutze, M. K., Strutt, F., Clarke, A. R., and Cameron, S. L. (2019). A transcriptome-based analytical workflow for identifying loci for species diagnosis: A case study with bactrocera fruit flies (Diptera: Tephritidae). *Aust. Entomol.* 58, 395–408. doi:10.1111/aen.12321

Kullander, S. O., Fang, F., Delling, B., and Ahlander, E. (1999). The fishes of the Kashmir valley, RiverJhelum, Kashmir valley: Impacts on the aquatic environment. *Swedmar Göteborg* 1999, 99–167.

Kumar, S. S., Gaur, A., and Shivaji, S. (2011). Phylogenetic studies in Indian scleractinian corals based on mitochondrial cytochrome b gene sequences. *Curr. Sci.* 101, 69–676.

Lakra, W. S., Verma, M. S., Goswami, M., Lal, K. K., Mohindra, V., Punia, P., et al. (2011). DNA barcoding Indian marine fishes. *Mol. Ecol. Resour.* 11, 60–71. doi:10.1111/j.1755-0998.2010.02894.x

Lal, K. K., Gupta, B. K., Punia, P., Mohindra, V., Saini, V. P., Dwivedi, A. K., et al. (2015). Revision of gonius subgroup of the genus Labeo Cuvier, 1816 and confirmation of species status of Labeo rajasthanicus (Cypriniformes: Cyprinidae) with designation of a neotype. *Indian J. Fish.* 62, 10e22.

Langer, S., Tripathi, N. K., and Khajuria, B. (2013). Morphometric and meristic study of golden mahseer (*Tor putitora)* from Jhajjar Stream (JandK), India. *Res. J. Animal, Veterinary Fish. Sci.* 1, 1–4.

Laskar, B. A., Bhattacharjee, M. J., Dhar, B., Mahadani, P., Kundu, S., and Ghosh, S. K. (2013). The species dilemma of northeast Indian mahseer (actinopterygii: Cyprinidae): DNA barcoding in clarifying the riddle. *PLoS One* 8, e53704–e53711. doi:10.1371/journal.pone.0053704

Laskar, B. A., Kumar, V., Kundu, S., Tyagi, K., and Chandra, K. (2018). Taxonomic quest: Validating two mahseer fishes (actinopterygii: Cyprinidae) through molecular and morphological data from biodiversity hotspots in India. *Hydrobiologia* 815 (1), 113–124. doi:10.1007/s10750-018-3555-6

Le, S., Josse, J., and $ Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* 25 (1), 1–18. doi:10.18637/jss.v025.i01

Li, Y., Ren, Z., Shedlock, A. M., Wu, J., Sang, L., Tersing, T., et al. (2013). High altitude adaptation of the schizothoracine fishes (Cyprinidae) revealed by the mitochondrial genome analyses. *Gene* 517, 169–178. doi:10.1016/j.gene.2012.12.096

Lohani, V., Pant, B., Pandey, N. N., and Ram, R. N. (2020). Morphological account of Schizothorax richardsonii population of river kosi and river Alaknanda. *Pharma Innov.* 9 (3), 30–33.

Lv, Y., Li, Y., Ruan, Z., Bian, C., You, X., Yang, J., et al. (2018). The complete mitochondrial genome of *Glyptothorax macromaculatus* provides a well-resolved molecular phylogeny of the Chinese sisorid catfishes. *Genes.* 9 (6), 282. doi:10.3390/genes9060282

Ma, Q., He, K., Wang, X., Jiang, J., Zhang, X., and Song, Z. (2020). Better resolution for Cytochrome b than Cytochrome c Oxidase subunit I to identify Schizothorax species (Teleostei: Cyprinidae) from the Tibetan Plateau and its adjacent area. *DNA Cell. Biol.* 39 (4), 579–598. doi:10.1089/dna.2019.5031

Maddison, D. R., and Maddison, W. P. (2003). *MacClade 4.06: Analysis of phylogeny and character evolution*. Sunderland, MA: Sinauer Associates.

Mardia, K. V., Kent, J., and Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.

Martinez, A. S., Willoughby, J. R., and Christie, M. R. (2018). Genetic diversity in fishes is influenced by habitat type and life-history variation. *Ecol. Evol.* 8, 12022–12031. doi:10.1002/ece3.4661

Mcclelland, J. (1839). Indian cyprinidae. *Asiat. Res.* 19, 217–471.

Mehmood, S., Ahmed, I., and Ali, N. (2021). Length-weight relationship, morphometric and meristic controlling elements of three freshwater fish species inhabiting North-Western Himalaya. *Egypt. J. Aquat. Biol. Fish.* 25 (6), 243–257. doi:10.21608/ejabf.2021.211325

Menon, A. G. K. (1964). *Monograph of the cyprinid fishes of the genus Garra Hamilton, Vol. 14*. Government of India.

Meyer, C. P., and Paulay, G. (2005). DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* 3, e422. doi:10.1371/journal.pbio.0030422

Mina, M. V., Levin, B. A., and Mironovsky, A. N. (2005). On the possibility of using character estimates obtained by different operators in morph metric studies of fish. *J. Ichthyol.* 45 (4), 284–294.

Mir, F. A., Mir, J. I., and Chandra, S. (2013a). Phenotypic variation in the snow trout *schizothoraxrichardsonii* (gray, 1832) (actinopterygii: Cypriniformes: Cyprinidae) from the IndianHimalayas. *Contrib. Zool.* 82, 115–122. doi:10.1163/18759866-08203001

Mir, F. A., Mir, J. I., Patiyal, R. S., and Chandra, S. (2013b). Pattern of morphometric differentiation among three populations of snowtrout, *Schizothorax plagiostomus* (Actinopterygii: Cypriniformes: Cyprinidae), from Kashmir Himalaya using a truss network system. *Acta Ichthyol. Piscat.* 43 (4), 277–284. doi:10.3750/aip2013.43.4.03

Misra, M., Nautiyal, P., and Lai, M. S. (1980). Morphological studies of hyobranchial skeleton of some Garhwal hill stream fishes. Hyobranchial skeleton of Schizothorx spp. *Indian J. Zootomy* 21 (1-3), 117–123.

Naeem, M., Salam, A., Ashraf, M., Khalid, M., and Ishtiaq, A. (2011). External morphometric study of hatchery reared mahseer (Tor putitora) in relation to body size and condition factor. *Afr. J. Biotechnol.* 10 (36), 7071–7077.

Negi, R. K., and Negi, T. (2010). Analysis of morphometric character of *Schizothorax richardsonii* (gray, 1832) from the uttarkashi district of uttrakhand state. *India. J. Biol. Sci.* 10 (6), 536–540.

Page, R. D. M., and Holmes, E. C. (1998). *Molecular evolution: A phylogenetic approach*. London: Blackwell Science.

Pandey, P. K., Singh, Y. S., Tripathy, P. S., Kumar, R., Abujam, S. K., and Parhi, J. (2020). DNA barcoding and phylogenetics of freshwater fish fauna of Ranganadi River, Arunachal Pradesh. *Gene* 754, 144860. doi:10.1016/j.gene.2020.144860

Pandey, S. K., and Nautiyal, P. (1997). Statistical evaluation of some meristic and morphometric characters of taxonomic significance in Schizothorax richardsonii (Gray) and Schizothorax plagiostomus (Heckel). *Indian J. Fish.* 44 (1), 75–79.

Pethiyagoda, R., and Kottelat, M. (1995). A review of the barbs of the Puntius filamentosus group (Teleostei: Cyprinidae) of southern India and Sri Lanka. *Raffles Bull. Zool. Suppl.* 12, 127–144.

Pollack, S. J., Kawalek, M. D., Williams-Hill, D. M., and Hellberg, R. S. (2018). Evaluation of DNA barcoding methodologies for the identification of fish species in cooked products. *Food control.* 84, 297–304. doi:10.1016/j.foodcont.2017.08.013

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25 (7), 1253–1256. doi:10.1093/molbev/msn083

Poulet, N., Reyjol, Y., Collier, H., Sovan, L., and Lek, S. (2005). Does fish scale morphology allow the identification of populations at a local scale? A case study for rostrum dace *Leuciscus leuciscus* burdigalensis in river viaur (SW France). *Aquat. Sci.* 67, 122–127. doi:10.1007/s00027-004-0772-z

Puillandre, N., Lambert, A., Brouillet, S., and Achaz, G. (2012). ABGD, automatic barcode gap Discovery for primary species delimitation. *Mol. Ecol.* 21 (8), 1864–1877. doi:10.1111/j.1365-294X.2011.05239.x

R Development Core Team (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rafique, M., and Khan, N. U. H. (2012). Distribution and status of significant freshwater fishes of Pakistan. *Rec. Zool. Surv. Pak.* 21, 90–95.

Rajpoot, A., Kumar, V. P., Bahuguna, A., and Kumar, D. (2016). DNA barcoding and traditional taxonomy: An integrative approach. *Int. J. Curr. Res.* 8, 42025–42031. doi:10.1139/gen-2015-0167

Rathipriya, A., Karal Marx, K., and Jeyashakila, R. (2019). Molecular identification and phylogenetic relationship of flying fishes of Tamil Nadu coast for fishery management purposes. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 30 (3), 500–510. doi:10.1080/24701394.2018.1558220

Regmi, B., Douglas, M. R., Edds, D. R., and Douglas, M. E. (2021). Geometric morphometric analyses define riverine and lacustrine species flocks of himalayan snowtrout (cyprinidae: Schizothorax) in Nepal. *Aquat. Biol.* 30, 19–31. doi:10.3354/ab00737

Rehman, A., Khan, M. F., Bibi, S., and Nouroz, F. (2020). Comparative phylogeny of (schizothoraxesocinus) with reference to 12s and 16S ribosomal RNA from River swat, Pakistan. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* 31 (2), 81–85. doi:10.1080/24701394.2020.1741561

Revelle, W. (2020). *psych: Procedures for personality and psychological research*. R package version 2.1.3. Evanston, Illinois, USA: Northwestern University. Available at: https://CRAN.R-project.org/package=psych.

Roul, S. K., Jeena, N. S., Kumar, R., Vinothkumar, R., Rahangdale, S., Rahuman, S., et al. (2021). Postulating the modality of integrative taxonomy in describing the cryptic congener pampus griseus (cuvier) and systematics of the genus pampus (perciformes: Stromateidae). *Front. Mar. Sci.* 8, 1–22. doi:10.3389/fmars.2021.778422

Sarkar, U. K., Pathak, A. K., Sinha, R. K., Sivakumar, K., Pandian, A. K., Pandey, A., et al. (2012). Freshwater fish biodiversity in the river ganga (India): Changing pattern, threats and conservation perspectives. *Rev. Fish. Biol. Fish.* 22, 251–272. doi:10.1007/s11160-011-9218-6

Sharma, J., Alka, P., Pratibha, B., and Jitendra, K. (2016). Assessment of genetic diversity of mahseer (*Tor tor*) using RAPD markers from wild and cultured conditions in Madhya Pradesh. *J. Exp. Zool. India* 19 (1), 23–29.

Sharma, N. K., Mir, J. I., Pandey, N. N., Akhtar, M. S., Bashir, A., and Singh, R. (2014). Meristic and morphometric characteristics of *Crossocheilus diplochilus* (heckel, 1838) from the Poonch valley of Jammu and Kashmir, India. *World J. Zoology* 9 (3), 184–189.

Shen, Y., Guan, L., Wang, D., and Gan, X. (2016). DNA barcoding and evaluation of genetic diversity in Cyprinidae fish in the midstream of the Yangtze River. *Ecol. Evol.* 6 (9), 2702–2713. doi:10.1002/ece3.2060

Sherzada, S., Khan, M. N., Babar, M. E., Idrees, M., Wajid, A., Sharif, M. N., et al. (2020). Identification of three cyprinidae family members through cytochrome oxidase I. *Pak. J. Zool.* 52 (1), 417–420. doi:10.17582/journal.pjz/2020.52.1.sc13

Silas, E. G. (1960). Fishes from the Kashmir valley. *J. Bombay Nat. Hist. Soc.* 57 (1), 66–77.

Simon, C., Franke, A., and Martin, A. (1991). "The polymerase chain reaction: DNA extraction and amplification," in *Molecular techniques in taxonomy*. Editors G. M. Hewitt, A. Johnson, and J. P. W. Young (Berlin: Springer), 329–355.

Talwar, P. K., and Jhingran, A. G. (1991). *Inland fisheries of India and adjacent countries.Vol.I& II*. New Delhi: Oxford & IBH Publishing Company, 1–1158.

Ter Braak, C. J. F., and Verdonschot, P. F. M. (1995). Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquat. Sci.* 57, 255–289. doi:10.1007/bf00877430

Thai, B. T., Si, V. N., Phan, P. D., and Austin, C. M. (2007). Phylogenetic evaluation of subfamily classification of the Cyprinidae focusing on Vietnamese species. *Aquat. Living Resour.* 20, 143–153. doi:10.1051/alr:2007025

Tsoupas, A., Papavasileiou, S., Minoudi, S., Gkagkavouzis, K., Petriki, O., Bobori, D., et al. (2022). DNA barcoding identification of Greek freshwater fishes. *PloS one* 17 (1), e0263118. doi:10.1371/journal.pone.0263118

Vishwanath, W., Lakra, W. S., and Sarkar, U. K. (2007). *Fish of northeast India*. Lucknow, India: National Bureau of Fish Genetic Resources.

Wahab, A., and Yousafzai, A. M. (2017). Quantitative attributes of family Sisoridae (Siluriformes) with a new record of *Glyptothorax kashmirensis* from River Panjkora, District Lower Dir, Khyber Pakhtunkhwa, PakistanThe BPP program for species tree estimation and species delimitation. *J. Entomol. Zool. Stud.* 5 (2), 741–745.

Wagle, S. K., Pradhan, N., and Shrestha, M. K. (2015). Morphological divergence of snow trout (Schizothorax richardsonii, Gray 1932) from rivers of Nepal with insights from a morphometric analysis. *Int. J. Appl. Sci. Biotechnol.* 3 (3), 464–473. doi:10.3126/ijasbt.v3i3.13123

Wani, I. F., Bhat, F. A., Balkhi, M. H., Shah, T. H., Shah, F. A., and Bhat, B. A. (2018). Study on Gonadosomatic index (GSI) during the three seasons (pre-spawning, spawning and post-spawning periods) of Schizothorax Niger Heckel in dal lake, Kashmir. *J. Pharmacogn. Phytochem.* 7 (6), 2131–2136.

Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., and Hebert, P. D. (2005). DNA barcoding Australia's fish species. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1847–1857. doi:10.1098/rstb.2005.1716

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Wickham, H., François, R., Henry, L., and Müller, K. (2022). *Dplyr: A grammar of data manipulation*. R package version 1.0.8. Available at: https://CRAN.R-project.org/package=dplyr.

Wu, Y. F., and Wu, C. Z. (1992). *The fishes of the qinghai-xizang plateau (in Chinese)*. Chengdu, China: Sichuan Publishing House of Science & Technology.

Xiao, W., Zhang, Y., and Liu, H. (2001). Molecular systematics of xenocyprinae (teleostei: Cyprinidae): Taxonomy, biogeography, and coevolution of a special group restricted in EastAsia. *Mol. Phylogenet. Evol.* 18, 163–173. doi:10.1006/mpev.2000.0879

Xu, L., Lin, Q., Xu, S., Gu, Y., Hou, J., Liu, Y., et al. (2018). Daphnia diversity on the Tibetan Plateau measured by DNA taxonomy. *Ecol. Evol.* 8 (10), 5069–5078. doi:10.1002/ece3.4071

Yan, G. J., He, X. K., Cao, Z. D., and Fu, S. J. (2013). An interspecific comparison between morphology and swimming performance in cyprinids. *J. Evol. Biol.* 26, 1802–1815. doi:10.1111/jeb.12182

Yang, J. Q., Tang, W. Q., Liao, T. Y., Sun, Y., Zhou, Z. C., Han, C. C., et al. (2012). Phylogeographical analysis on Squalidusargentatus recapitulates historical landscapes and drainage evolution on the island of Taiwan and mainland China. *Int. J. Mol. Sci.* 13, 1405–1425. doi:10.3390/ijms13021405

Yang, L., Sado, T., Hirt, M. V., Pasco-Viel, E., Arunachalam, M., Li, J., et al. (2015). Phylogeny and polyploidy: Resolving the classification of cyprinine fishes (Teleostei: Cypriniformes). *Mol. Phylogenet. Evol.* 85, 97–116. doi:10.1016/j.ympev.2015.01.014

Yang, Z. (2015). The BPP program for species tree estimation and species delimitation. *Curr. Zool.* 61 (5), 854–865. doi:10.1093/czoolo/61.5.854

Zhuang, X., Qu, M., Zhang, Z., and Ding, S. (2013). A comprehensive description and evolutionary analysis of 22 grouper (Perciformes, Epinephelidae) mitochondrial genomes with emphasis on two novel genome organizations. *PLoS One* 8, e73561. doi:10.1371/journal.pone.0073561

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fgene.2023.1029200



OPEN ACCESS

EDITED BY
Yuzine Esa,
Putra Malaysia University, Malaysia

REVIEWED BY
Hong Wei Liang,
Yangtze River Fisheries Research Institute (CAFS), China
Rex Dunham,
Auburn University, United States

*CORRESPONDENCE
Hong-Yi Gong,
hygong@mail.ntou.edu.tw

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

RECEIVED 26 August 2022
ACCEPTED 02 January 2023
PUBLISHED 13 January 2023

CITATION
Chu W-K, Huang S-C, Chang C-F, Wu J-L and Gong H-Y (2023), Infertility control of transgenic fluorescent zebrafish with targeted mutagenesis of the *dnd1* gene by CRISPR/Cas9 genome editing.
*Front. Genet.* 14:1029200.
doi: 10.3389/fgene.2023.1029200



# Infertility control of transgenic fluorescent zebrafish with targeted mutagenesis of the *dnd1* gene by CRISPR/Cas9 genome editing

Wai-Kwan Chu[1,2], Shih-Chin Huang[1], Ching-Fong Chang[1,2], Jen-Leih Wu[3,4] and Hong-Yi Gong[1,2]*

[1]Department of Aquaculture, National Taiwan Ocean University, Keelung, Taiwan, [2]Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan, [3]Institute of Cellular and Organismic Biology, Academia Sinica, Taipei, Taiwan, [4]College of Life Sciences, National Taiwan Ocean University, Keelung, Taiwan

Transgenic technology and selective breeding have great potential for the genetic breeding in both edible fish and ornamental fish. The development of infertility control technologies in transgenic fish and farmed fish is the critical issue to prevent the gene flow with wild relatives. In this study, we report the genome editing of the *dead end* (*dnd1*) gene in the zebrafish model, using the CRISPR/Cas9 technology to achieve a loss-of-function mutation in both wild-type zebrafish and transgenic fluorescent zebrafish to develop complete infertility control technology of farmed fish and transgenic fish. We effectively performed targeted mutagenesis in the *dnd1* gene of zebrafish with a single gRNA, which resulted in a small deletion (–7 bp) or insertion (+41 bp) in exon 2, leading to a null mutation. Heterozygotes and homozygotes of *dnd1*-knockout zebrafish were both selected by genotyping in the $F_1$ and $F_2$ generations. Based on a comparison of histological sections of the gonads between wild-type, heterozygous, and homozygous *dnd1* zebrafish mutants, the *dnd1* homozygous mutation (aa) resulted in the loss of germ cells. Still, there was no difference between the wild-type (AA) and *dnd1* heterozygous (Aa) zebrafish. The homozygous *dnd1* mutants of adult zebrafish and transgenic fluorescent zebrafish became all male, which had normal courtship behavior to induce wild-type female zebrafish spawning. However, they both had no sperm to fertilize the spawned eggs from wild-type females. Thus, all the unfertilized eggs died within 10 h. The targeted mutagenesis of the *dnd1* gene using the CRISPR/Cas9 technology is stably heritable by crossing of fertile heterozygous mutants to obtain sterile homozygous mutants. It can be applied in the infertility control of transgenic fluorescent fish and genetically improved farmed fish by selective breeding to promote ecologically responsible aquaculture.

KEYWORDS

**infertility control, fluorescent zebrafish, transgenic fish, genome editing, CRISPR/Cas9, Dead end, PGCs**

## 1 Introduction

Transgenic technology is a powerful and useful technique to establish transgenic organisms by insertion of a foreign gene into the genome of organisms to obtain or enhance the function of transgene *in vivo* under the control of ubiquitous promoter or tissue-specific promoter (Ozato et al., 1992; Gordon, 1997; Her et al., 2003). Since 1980, scholars have used transgenic technology to improve the performance of major farmed fish species (Du et al., 1992; Maclean, 1993). A transgenic Atlantic salmon, AquAdvantage Salmon, which expressing

growth hormone of Chinook salmon was established under the control of a promoter from ocean pout and obtained the approval for consumption in the U.S. on November 2015 and 6-month later in Canada. AquAdvantage Salmon grows twice faster to market size than non-transgenic Atlantic salmon counterpart. Thus, transgenic technology has great potential for the development of aquaculture, as it significantly reduces farming costs and risks (Cook et al., 2000; Muir, 2004). With the subsequent simplification of the transgenic procedures and methods, the application of transgenic technology gradually became popular (Kawakami, 2007; Suster et al., 2009). Besides in the food fish aquaculture industry, transgenic technology has been applied in the ornamental industry (Gong et al., 2003). Fluorescent fish is a concept originated by scholars using the transgenic technique to visualize and calibrate tissues to study specific gene regulatory mechanisms and functions (Ju et al., 1999; Gong et al., 2001; Chudakov et al., 2010). According to the ornamental fish industry data, fluorescent fish are new varieties that meet consumers' demands in terms of unique body colors, posing a huge potential business opportunity, which may enable the industry to become more competitive (Stewart, 2006). Transgenic fish applications are expected to increase exponentially in both the aquatic and ornamental fish industries globally. If these transgenic fish escape and interbreed with wild stock, this poses a potential threat to the ecosystem and environment (Hu et al., 2007). Recently, transgenic glowing zebrafish appear to be thriving after escape from fish farms into Brazilian streams and may threaten local biodiversity (Magalhaes et al., 2022; Moutinho, 2022). We believe that developing practical infertility control technology is the most effective way to avoid ecological risks, and it also promotes the development of an environmentally responsible aquaculture (Muir and Howard, 1999; Stigebrandt et al., 2004; Wong and Van Eenennaam, 2008).

The method of establishing infertile individuals has been continuously developed. The methods commonly used to produce infertile fish are triploidization and interspecies hybridization (Benfey et al., 1989; Arai, 2001; Cal et al., 2006); however, some treated individuals were found to maintain fertility and show limitations, depending on the species (Wagner et al., 2006; Piferrer et al., 2009). In our previous research, we tried to use the transgenic strategy, transferring the germ cells' specific gene *piwi* promoter to be combined with the nitroreductase toxic protein (NTR). When the metronidazole (Mtz) bath immersion treatment succeeded, cell apoptosis began due to the conversion of Mtz into a cytotoxic compound, which may affect the germ cells and induce gonadal development disruptions. However, the effects of the NTR/Mtz system are not obvious. While the methods mentioned above failed, the disruption of the mechanisms of primordial germ cell formation and migration was successful in achieving complete infertility (Wong and Collodi, 2013; Wong and Zohar, 2015a; Wong and Zohar, 2015b).

Primordial germ cells (PGCs) are the progenitor cells of gametes. PGCs play an important role, as egg and sperm production relies on the formation, differentiation, and correct localization of PGCs (Yoon et al., 1997; Weidinger et al., 1999; Xu et al., 2010; Yamaha et al., 2010). In zebrafish, the development of PGCs has been studied extensively. PGC formation occurs during early embryogenesis maternally through a germplasm composed of maternal RNA-binding proteins and mRNAs (Braat et al., 1999; Extavour and Akam, 2003; Raz, 2003). In addition, the germ cell marker *ddx4* also named as *vasa*, an evolutionarily conserved gene, was first discovered to label PGCs successfully, and it was shown to be a key factor in the migration of PGCs to the genital ridge (Yoon et al., 1997; Braat et al., 1999; Castrillon et al., 2000; Yoshizaki et al., 2000). However, by inhibiting the gene expression by gene silencing in different species, it did not induce a state of a complete loss of germ cells or infertility, resulting in a single sex (Raz, 2000; Li et al., 2009; Hartung et al., 2014). There is no evidence proving that no other cells can potentially contribute to the germline during normal development or when the number of *ddx4*-expressing cells is reduced. The *dead end (dnd1)* gene, encoding an RNA-binding protein DND1, was first identified in zebrafish (Weidinger et al., 2003) and is highly conserved in vertebrate species (Horvay et al., 2006; Aramaki et al., 2007; Liu et al., 2009; Nagasawa et al., 2013). The expression of some germline-specific mRNAs relied on the protection of DND1 from microRNA-mediated inhibition (Kedde et al., 2007a). The DND1 knockdown by translational inhibition of maternal *dnd1* mRNA with Morpholino antisense oligonucleotides resulted in PGCs losing their ability to migrate actively and mis-migration, followed by apoptosis or transdifferentiation into somatic cells in zebrafish (Weidinger et al., 2003; Gross-Thebing et al., 2017). The *dead end* (*dnd1*) knockout fish by genome editing was first conducted in medaka with TALENs and showed germ cell-less gonads (Wang and Hong, 2014; Wang, 2015). The *dnd1* knockout zebrafish was first established by genome editing with Zinc Finger Nucleases (ZFNs) as germ cells-less recipients of surrogate for germ cell transplantation (Li et al., 2017). However, compared to ZFN and TALEN, which rely on DNA binding domain to recognize DNA, the CRISPR/Cas9 system by using a single guide RNA (sgRNA) for DNA recognition is more efficient, convenient, and cost-effective (Doudna and Charpentier, 2014). Therefore, CRISPR/Cas9 technology had been applied in genome editing of several fish species such as Atlantic salmon, medaka, sterlet, and rainbow trout to target *dnd1* to achieve infertility (Wargelius et al., 2016; Sawamura et al., 2017; Baloch et al., 2019; Baloch et al., 2021; Fujihara et al., 2022).

CRISPR/Cas9 technology has been widely used due to its ease of operation, which allows scholars to investigate gene function (Ran et al., 2013). Genome editing used the Cas9 protein and guide RNA (gRNA), and the 5' of gRNA contained an 18–20 bp protospacer sequence that could identify the target sequence and guide Cas9 to cause a DNA double-strand break, thus inducing the mechanism of DNA repair, i.e., non-homologous end joining (NHEJ) and homology-directed repair (HDR) (Garneau et al., 2010; Horvath and Barrangou, 2010; Hwang et al., 2013). There is an opportunity to cause an insertion or deletion through the NHEJ repair mechanism, and an in-frame stop codon may emerge through the insertion or deletion performed by CRISPR/Cas9, causing the translation to stop early, which results in a loss of gene function due to early truncation of the protein (Chapman et al., 2012; Auer et al., 2014).

While transgenic technology has a huge potential to be applied in aquaculture, it poses unexpected risks to the ecosystem, as the organisms maintain the ability to crossbreed with wild stock. Developing effective and practical infertility control is the key to solving the problem. Although *dnd1* knockout in zebrafish was achieved by ZFNs (Li et al., 2017), CRISPR/Cas9 technology was still not applied in *dnd1* knockout of zebrafish, especially in new application of infertility control of transgenic fluorescent zebrafish as a model of fluorescent ornamental fish. The phenotypes of homozygous and heterozygous *dnd1* knockout zebrafish by CRISPR/Cas like those by ZFNs are not novel but convince us of no off-targeting effect in

*dnd1* knockout zebrafish by both genome editing technologies. In this study, we demonstrate that the *dnd1* knockout by CRISPR/Cas9 genome editing had been successfully applied to the infertility control of transgenic fluorescent zebrafish as a model of transgenic fluorescent ornamental fish to prevent potential impact on ecology when they escape into the wild field.

## 2 Materials and methods

### 2.1 Establishment of *dnd1*-knockout zebrafish using the CRISPR/Cas9 system

Wild-type (AB) strain zebrafish were used as experimental animals in this study. The zebrafish were maintained in a recirculating system, with the temperature maintained at 28°C and photoperiodism of 14 h of light and 10 h of dark. The animal use protocol of this research had been reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of National Taiwan Ocean University. The IACUC Approval No. is 108042. The CRISPR/Cas9 system was used to establish the *dnd1*-knockout zebrafish line (Baloch et al., 2021). The sgRNA target site was designed using CHOPCHOP, using an off-target assay for analysis (Labun et al., 2016). The pCS2-nCas9n was a gift from Wenbiao Chen (Addgene plasmid # 47929; http://n2t.net/addgene:47929; RRID: Addgene_47929) for Cas9 mRNA synthesis by *in vitro* transcription (Jao et al., 2013). The zebrafish *dnd1* target-specific DNA oligo (5′-TTCTAATACGACTCACTATAGTAACCCAAG TCAATGGGCAGGTTTTAGAGCTAGA-3′) was synthesized by "Genomics" company (Taiwan) for sgRNA synthesis. The sgRNA was synthesized based on the protocol of EnGen® sgRNA synthesis kit (New England Biolabs, United States). The zebrafish reproduction was performed under the conditions mentioned above. The embryos were collected to perform the microinjection. The microinjection was performed by co-injecting a mixture of reagents—300 ng/μL of Cas9 mRNA and 30 ng/μL of sgRNA—into one-cell-stage embryos. The injected embryos were cultivated in a Petri dish at 28°C for hatching. To predict the efficiency of the gRNA targeting, 10 individuals were randomly selected 24 h after fertilizing (hpf) the injected embryos for DNA extraction, and the T7 endonuclease I (T7E1) assay was used to evaluate the efficiency of the gRNA target site mutation (Urnov et al., 2005).

### 2.2 Mutation screening by sequencing in second-generation ($F_1$) zebrafish

The $F_1$ generation was produced by crossing the first-generation ($F_0$) with wild-type and inbreeding of the first-generation ($F_0$). Genomic DNA from the caudal fin of each $F_1$ fish larva was extracted to analyze the genotype (Meeker et al., 2007). A 150 bp fragment of the target site region was amplified using the 5× PCR Dye Hot Start Master Mix (GeneMark, Taiwan), using the following primers: 5′-ATTCTGAACCCGCAGAAACTCAAGTCTCTGCAG GA and 3′-AATGCTGCTCTCAGGTCGATGGAGGGGCACTTA C. The genotype analysis was performed by gel electrophoresis using a 2% agarose gel, followed by the gene cloning of the individuals to confirm the mutations. The sequences were analyzed using the Molecular Evolutionary Genetics Analysis Version 7.0 (Mega 7) software to predict the form of protein (Kumar et al., 2016; Bhattacharya and Van Meir, 2019). Heterozygotes whose mutations were predicted to show a stop codon were selected to develop the $F_1$ generation. The same mutation pattern of heterozygotes in the $F_1$ generation was inbred to establish the $F_2$ generation. The genotyping process mentioned above was performed to interpret the *dnd1*-knockout heterozygotes and homozygotes, and the number of individuals of different genotypes was recorded to calculate the heritability.

### 2.3 Establishment of the fluorescent *dnd1*-knockout zebrafish line

The transgenic fluorescent zebrafish line *Tg(-2.4ckmb:TcCFP13)*, expressing the Taiwan coral (*Acropora* sp.) cyan fluorescent protein, TcCFP-13 cDNA (provided by Dr. Ming-Chyuan Chen, National Kaohsiung University of Science and Technology, Taiwan), driven by a novel zebrafish muscle-specific *ckmb* 2.4 kb promoter/enhancer (GenBank accession number HM347596), was previously established *via* the *Tol2* transposon system (Gong et al., 2015). The fluorescent zebrafish strain was mated with the *dnd1*-knockout heterozygotes to establish the $F_1$ generation fluorescent *dnd1*-knockout fish line. The individuals with cyan fluorescent embryos were selected after 72 hpf, followed by the genotype analysis process mentioned above to select the heterozygotes. The cyan fluorescent *dnd1*-knockout homozygous zebrafish were obtained by inbreeding with the $F_1$ *dnd1*-knockout heterozygotes.

### 2.4 PGC localization analysis by whole-mount *in situ* hybridization

The sense and antisense *ddx4* riboprobes were designed in the open reading frame of zebrafish *ddx4*, and the primers of the probe were 5′-GCGTGTCCACCTGCTACCGGCTCTTCTGAA and 3′-TTCATCACGGGAGCCACTGCGAAAACC. The riboprobes were synthesized using the DIG RNA Labeling Kit (SP6/T7) (Roche) from linearized pGEM-T *ddx4*. The embryos were collected 24 h post-fertilization, followed by fixation with 4% paraformaldehyde. The signals were detected by NBT/BCIP staining (Brend and Holley, 2009). We selected and separated the embryos with a positive and negative signal under a Leica EZ4 microscope (Leica, Germany), after capturing the phenotypes. Then, the embryos were washed 5 times using 1× PBS, followed by DNA extraction with the MasterPure™ DNA Purification Kit (Epicentre, United States). PCR amplification was performed using the 5× PCR Dye Hot Start Master Mix (GeneMark, Taiwan). The genotype was analyzed by gel electrophoresis using a 2% agarose gel.

### 2.5 Histological analysis of zebrafish gonadal tissue

Three individuals from the wild types, heterozygotes, and homozygotes in the female and male groups were selected. The sample (whole fish) was fixed with Davidson's fixation solution for 24 h; then, the sample was kept in 70% ethanol (Miki et al., 2018). The sample was dehydrated with ethanol and then embedded into paraffin

blocks. Then, the sample was cut into 5 μm-thick sections, and the sections were mounted on glass slides in a 42°C thermostatic water bath. Staining was performed using hematoxylin and eosin (HE stains), and the sections were analyzed using an optical microscope (Blazer, 2002).

## 2.6 Semi-quantitative PCR analysis of the gene expression level

Three individuals from the wild types, heterozygotes in the female and male groups, and homozygotes were selected for sampling of the gonadal tissues. The total RNA of the gonads was extracted using Ambion™ TRIzol Reagent and purified with the PureLink™ RNA mini kit. Reverse transcription was performed to synthesize the corresponding cDNA. The primers used in this study were as follows: *ddx4* (forward: 5′-ATGGATGACTGGGAGGAAGATCAG AGTCCCG-3’; reverse: 5′- TTCCATTTTCATCATTTTCATCAC GGGA-3′); *dazl* (forward: 5′-ATGGTTCAGGGGG TTCAGTTACCCGTGT-3’; reverse 5′-TGATGGTGGGGCCAG GCCTGGAGGACAGCA-3′); *nanos3* (forward: 5′-TTCTGGAAT GACTATCTCGGCCTGTCCA-3’; reverse: 5′-ATTGTGTGCGCG TTGTCCCCGTTGGCACTGC-3′); *tdrd7a*: (forward: 5′-ATGGCG GACGAGGAACTGGTGAAGAA-3’; reverse: 5′-CCGCGACCT CCACCTCGCCCCAGACGTCCAC GG-3′); *amh* (forward: 5′-ATGCTTTTCCAGGCAAGATTTGGGCTGATG-3’; reverse 5′-ACTGTCTCCTTTAACAGGATTGACACTGAACAG-3′); and *eef1α1l1* (forward: 5′-TGCCTTCG TCCCAATTTCAG-3’; reverse: 5′-CTCATGTCACGCACACAGCAAAG-3′). Semi-quantitative PCR was performed using the following procedure: 94°C for 2 min; 94°C for 30 s; 60°C for 30 s; 72°C for 1 min per kb (repeated for 25–35 cycles); 72°C for 5 min; the reaction was stopped at 4°C. The semi-quantitative PCR analysis was performed by gel electrophoresis using a 1% agarose gel.

## 2.7 Courtship behavior analysis of *dnd1*-knockout zebrafish

The *dnd1*-knockout homozygotic individuals were selected and crossbred with wild-type female zebrafish randomly. The wild-type groups were selected randomly as a control. The photoperiodism was 14 h light and 10 h dark, and the observation was performed during breeding, collecting the eggs of every group and counting and recording photos every hour. The same process was used for the cyan fluorescent *dnd1*-knockout homozygotic groups.

# 3 Results

## 3.1 Establishment of the *dnd1*-knockout zebrafish line

### 3.1.1 Targeted mutagenesis of the *dnd1* gene in zebrafish using CRISPR/Cas9

According to previous studies on the *dnd1* gene in zebrafish, the translation region that starts to form the key functional domain, the RNA-recognition motif (RRM), is located in exon 3. The target site of a single gRNA was designed to locate in the second exon of the *dnd1* gene in zebrafish to disrupt the formation of the RNA-recognition motif (Figure 1A). The components of synthesized gRNA and Cas9 mRNA were injected into one cell of fertilized zebrafish embryos (Figure 1B). The injected embryos were cultivated in Petri dishes, and at 24 h post-fertilization (hpf), we randomly selected ten embryos to investigate the efficiency of the target site mutation generated by gRNA using the T7E1 assay. According to the T7E1 results, in nine out of 10 embryos, multiple banding patterns were observed, with some individuals showing significant second major band patterns, revealing DNA mismatches formed in a fraction of heteroduplexes and cleaved by T7E1 (Figure 1C). It was confirmed that the mutation was constituted through the gRNA, and *dnd1* may be successfully disrupted by the CRISPR/Cas9 genome editing system in zebrafish. The $F_0$ generation was maintained until sexual maturity; then, the knockout effect was evaluated by extracting the DNA from the caudal fin, followed by PCR and gel electrophoresis to select the individuals. We predict a higher genetic efficiency for heritable mutations, which appeared in multiple bands (Figure 1D). Next, we conducted the inbreeding of selected F0 individuals to establish the $F_1$ generation.

### 3.1.2 Inheritable mutation analysis of $F_1$ generation

We further analyzed the mutation patterns of the $F_1$ generation. The fish larvae were cultivated for 1 month, and the caudal fin was cleaved for DNA extraction, following by PCR amplification for every individual. In comparing the bands of the PCR product of the wild-type, the individuals with the extra band were separated. To investigate whether the unexpected insertion or deletion mutation patterns in the coding sequence could generate an in-frame stop codon, we performed cloning and used the MEGA 7 software to conduct sequencing. In the $F_1$ generation, a total of 120 $F_1$ individuals were screened and analyzed. We observed several mutation patterns, five of which had effects on the coding sequence, and the obtained sequence patterns were deletions of seven base pairs (bp) in different regions and insertions of 41 bp (Figure 2A). As the patterns mentioned above resulted in an in-frame shift of the coding sequence, the stop codon was predicted to appear in the protein translation process. We classified the $F_1$ zebrafish mutation patterns as a, b, c, d, e, and f. In total, 10 out of 120 individuals were heterozygotes (c, d, e, and f), and two were homozygotes. Besides the heterozygotes, we only obtained two homozygous mutant $F_1$ individuals in the a and b patterns by inbreeding the $F_0$ generation. Considering the genetic background of $F_1$ generation may be affected by the inbreeding of $F_0$, we crossed the F0 zebrafish with wild-type to generate the c, d, e, and f strains to eliminate the noise of genetic background for this study. Pattern b, in particular, shows two different mutant alleles (b-1: +41 bp and b-2: −7 bp) (Figure 2B). To generate the stable heritage of the *dnd1*-knockout zebrafish fish line, subsequent mating was performed using individuals identified as heterozygotes patterns c and d (as d, e, and f’s predicted amino acid sequences were identical). In the $F_2$ generation, three genotypes were obtained: wild-type (AA, $dnd1^{+/+}$), heterozygotes (Aa, $dnd1^{+/-}$), and homozygotes with a mutation in a pair of alleles (aa, $dnd1^{-/-}$). To investigate the preliminary data of heterozygotes and homozygotes that might have been produced after inbreeding, two groups of $F_2$ offspring were subjected to genotyping 30 days post-fertilization. The results for the genotypes’ heritability ratios were similar for the two groups. In the mutation pattern c fish line, among the total of

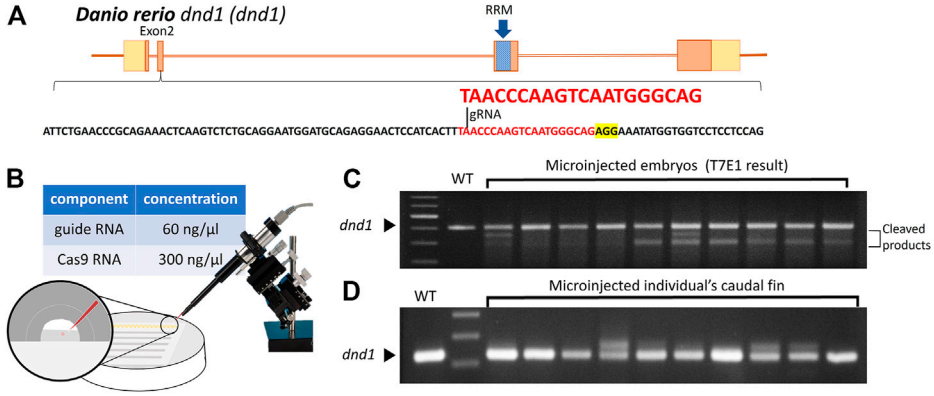
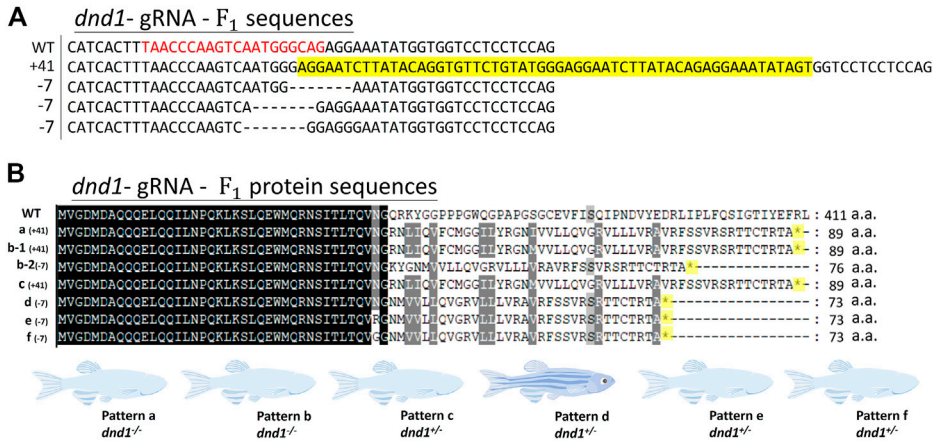



FIGURE 1
**(A)** Design of the CRISPR/Cas9 target sites of the guide RNA (gRNA) in *dnd1*. The gRNA was designed in exon 2, and the target site of the gRNA is indicated in red, while the sequence highlighted in yellow is PAM. The RNA-recognition motif (RRM) is the conserved functional domain encoded by exon 3 of the *dnd1* gene. **(B)** Microinjection schematic with the component and concentration of the working solution. **(C)** Mutation identified in each $F_0$ generation by T7E1 using 2% agarose gel. **(D)** Mutation efficiency analyzed using PCR on the caudal fin's DNA.




| $F_1$ inbreeding groups | $F_2$ embryos sequences analyses | | |
|---|---|---|---|
| | *dnd1*$^{+/+}$/Total | *dnd1*$^{+/-}$/Total | *dnd1*$^{-/-}$/Total |
| c (+41/WT) | 39/138 (28%) | 68/138 (49%) | 31/138 (23%) |
| d (-7/WT) | 23/114 (20%) | 62/114 (54%) | 29/114 (26%) |

FIGURE 2
**(A)** Mutation sequences identified in the $F_1$ generation by sequencing. The analyzed individuals are shown in order of the mutation sequence size of the insertions (+)/deletions (–). The target site of gRNA is indicated by red characters, and the yellow highlighted sequences are insertional sequences. **(B)** Mutation amino acid sequences identified in each $F_1$ zebrafish by genotyping. The in-frame stop codon is shown as * and highlighted in yellow. **(C)** Analysis of the ratio of different genotypes in $F_2$ generation.

138 individuals that we subjected to genotyping, 68 individuals (49%) were found to be *dnd1*$^{+/-}$ (Aa), which is the most populous group, 39 individuals (28%) were found to be *dnd1*$^{+/+}$ (AA), and 31 individuals (23%) were found to be *dnd1*$^{-/-}$ (aa) (Figure 2C). In the mutation pattern d fish line, 62 individuals (54%) found to be *dnd1*$^{+/-}$ (Aa) accounted for the highest proportion of the population among 114 individuals, 29 individuals (26%) were found to be *dnd1*$^{-/-}$ (aa), and 23 individuals (20%) were found to be *dnd1*$^{+/+}$ (AA) (Figure 2C). Considering the definiteness of the following experiments, we decided to further analyze the stain of pattern D with a heritable 7 bp deletion.

## 3.2 PGC and gonad analyses of *dnd1* zebrafish mutants

### 3.2.1 PGC localization analysis of the *dnd1*-knockout zebrafish larva

In wild-type zebrafish, PGCs may be observed by whole-mount *in situ* hybridization in 24 hpf embryos localized on the genital ridge. We performed the inbreeding of pattern D heterozygotes to collect the fertilized embryos. To investigate whether the PGCs localized normally on the genital ridge in *dnd1*-knockout offspring, the germ cell marker *ddx4* was used as an RNA probe label for PGC

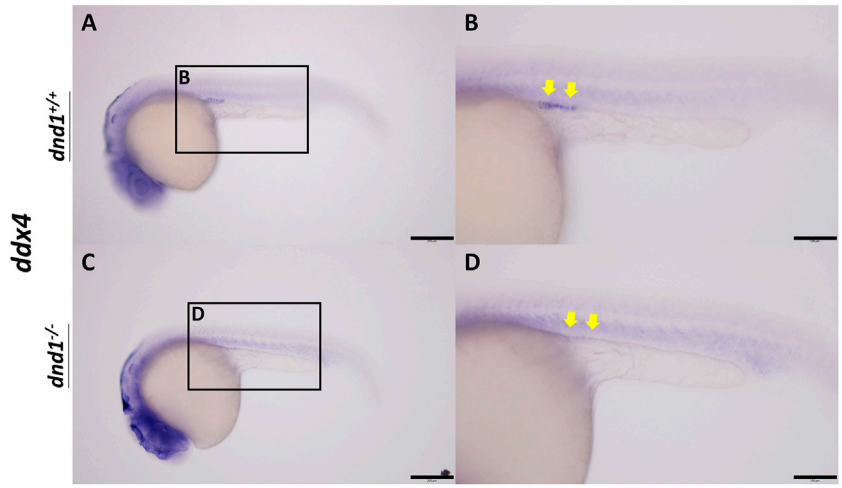



FIGURE 3
*ddx4* expression visualized by whole-mount *in situ* hybridization in 24-h post-fertilized embryos of the $F_2$ generation. **(A)-(B)** Wild-type zebrafish embryo. **(C)-(D)** Mutant zebrafish embryo. Bars = 200 μm **(A, C)** and 100 μm **(B, D)**.

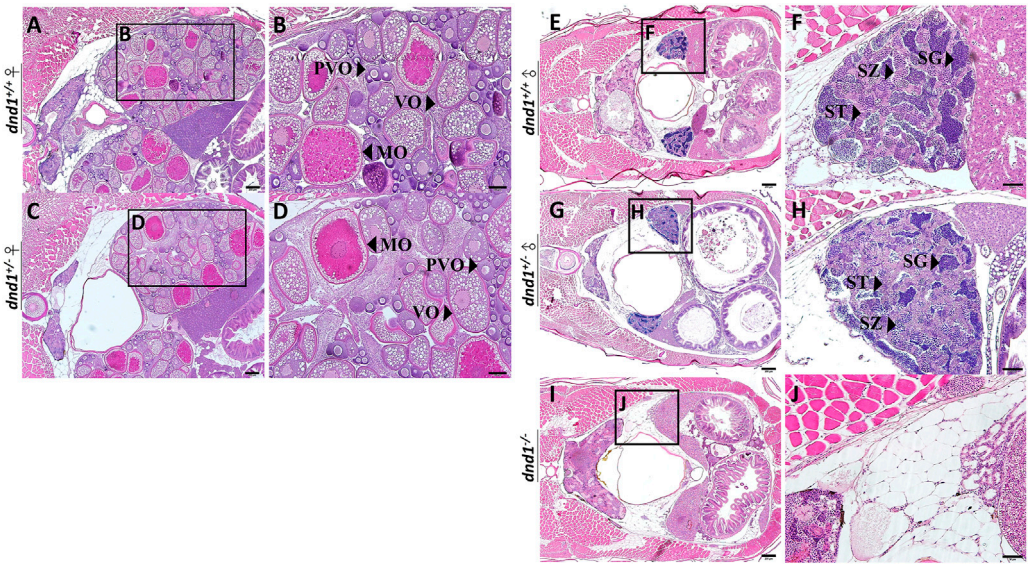



FIGURE 4
Histological analyses of zebrafish gonad. **(A)-(B)** Wild-type zebrafish ovary. **(C)-(D)** Heterozygote zebrafish ovary. **(E)-(F)** Wild-type zebrafish testis. **(G)-(H)** Heterozygote zebrafish testis. **(I)-(J)** Mutant zebrafish testis. PVO: previtellogenic oocytes; VO: vitellogenic oocytes; MO: mature oocyte; SC: spermatocytes; SG: spermatogonia; ST: spermatids. Bars = 200 μm **(A, C, E, G, I)**, 100 μm **(B, D)** and 50 μm **(F, H, J)**.

visualization. In the *dnd1*$^{+/+}$ (AA) individuals, we observed cell-like positive signals in dotted particles on the lateral view, localized in position straight in front of the tube-like region of the embryos (Figures 3A, B). Conversely, no positive-signal cells were found to be localized in the same position as the *dnd1*$^{-/-}$ (aa) embryos (Figures 3C, D). As *ddx4* is specifically expressed in PGCs, starting from the earliest stage of embryonic development, a loss of *ddx4* signal in mutant individuals represents loss of PGCs. Furthermore, we focused on the development of gonads in sexually mature individuals to discover whether *dnd1* knockout led to infertility.

### 3.2.2 Gonadal tissue analysis of the mature *dnd1*-knockout zebrafish

We separated the individuals according to the three groups of genotypes and randomly selected both male and female individuals from each genotype. No individual with the female phenotype was obtained from the *dnd1*$^{-/-}$ (aa) groups. Histological analysis was performed by tissue sectioning, and the tissue was stained with hematoxylin and eosin (HE staining). In the *dnd1*$^{+/+}$ (AA) female individuals, we observed a completely developed ovary, with primary, secondary, and mature oocytes (Figures 4A, B). In the *dnd1*$^{+/-}$ (Aa)

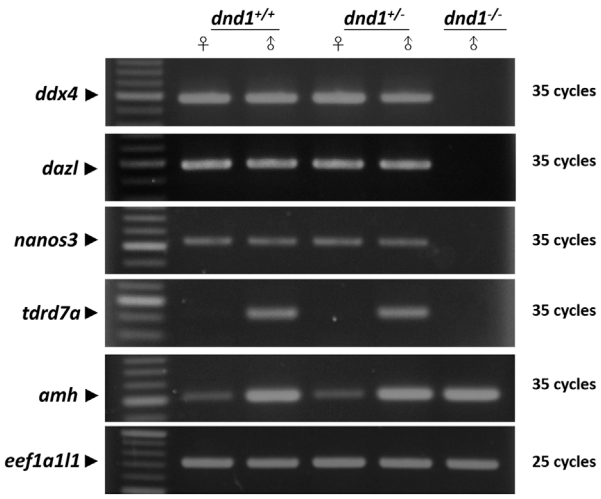



FIGURE 5
Analysis of the expression of germ-cell-related genes between the wild-type (*dnd1*$^{+/+}$), heterozygous (*dnd1*$^{+/-}$), and homozygous (*dnd1*$^{-/-}$) mutant zebrafish through semi-quantitative RT-PCR.

female individuals, there seemed to be no difference from the *dnd1*$^{+/+}$ (AA) heterozygote females, which maintained a developed ovary structure (Figures 4C, D).

In the male individual groups, we observed that the *dnd1*$^{+/+}$ (AA) individuals maintained a well-developed testis structure, with spermatogonia, spermatids, and spermatozoa (Figures 4E, F). Similarly, the *dnd1*$^{+/-}$ (Aa) male individuals had completely developed testes, which were not different from those of the *dnd1*$^{+/+}$ (AA) (Figures 4G, H). The all-male phenomenon was observed in the *dnd1*$^{-/-}$ (aa) individuals, compared with the *dnd1*$^{+/+}$ (AA) and *dnd1*$^{+/-}$ (Aa) individuals, and in the mutant of the homozygous knockout mutant (aa) groups, the *dnd1*$^{-/-}$ (aa) zebrafish completely lost their germ cells, and the left space of the tube-like testes was occupied by adipocytes (Figures 4I, J).

To determine whether the *dnd1*$^{+/+}$ (AA), *dnd1*$^{+/-}$ (Aa), and *dnd1*$^{-/-}$ (aa) knockouts would affect the expression levels of the germline-related genes *ddx4, dazl, nanos3,* and *tdrd7a,* the gonadal tissue of the testes and ovaries from *dnd1*$^{+/+}$ (AA) and *dnd1*$^{+/-}$ (Aa) and tube-like gonads from *dnd1*$^{-/-}$ (aa) were sampled for RNA extraction for analysis through semi-quantitative RT-PCR. In this experiment, the *anti-mullerian hormone* (*amh*) was used as a positive control, and *eef1α1l1* was the stable reference gene. In both the dnd1$^{+/+}$ (AA) and *dnd1*$^{+/-}$ (Aa) female and male groups, *ddx4, dazl,* and *nanos3* had similar expression levels, but in *dnd1*$^{-/-}$ (aa), the germ-cell-related gene expression was found to be lacking (Figure 5). While *dnd1*$^{+/+}$ (AA) and *dnd1*$^{+/-}$ (Aa) showed a specific expression of *tdrd7a* in the testes, *dnd1*$^{-/-}$ (aa) also showed a lack of expression (Figure 5). The gene *amh* showed a positive expression in both sexes of *dnd1*$^{+/+}$ (AA), *dnd1*$^{+/-}$ (Aa), and *dnd1*$^{-/-}$ (aa) (Figure 5). As *amh* is expressed in somatic cells of yolk-forming oocytes and spermatogonia gonad tissue, in the three groups of genotypes, males had a higher expression level than the females, but no difference was observed in *amh* expression between the same sexes for the three genotypes (Figure 5) by quantitation of *amh* bands with ImageJ analysis and normalized with *eef1α1l*. The *amh* in *dnd1*$^{-/-}$ (aa) showed comparable expression level with male fish of *dnd1*$^{+/+}$ (AA), and *dnd1*$^{+/-}$ (Aa). This indicates that the germ-cell-specific gene expression of the *dnd1*$^{-/-}$ (aa) individual was lost, but it also showed that gene expression in the gonad somatic cells was maintained.

### 3.2.3 Sexual characteristics and courtship behavior analysis

We analyzed the mating behavior and fertilization capabilities of the *dnd1*-knockout zebrafish. The results show that *dnd1*$^{-/-}$ (aa) are male and sustain a normal courtship behavior, which allows the wild-type female zebrafish to spawn eggs (Figure 6I). The spawned eggs were collected and observed under a dissecting microscope. Normally, fertilized zebrafish embryos reach a 50% epiboly stage after 6 h (Figures 6IIA, B). The eggs that are spawned by *dnd1*$^{-/-}$ (aa) and wild-type female zebrafish show half of the animal pole and half of the vegetal pole in terms of their shape. This is a phenomenon that indicates that eggs are not fertilized (Figures 6IIC, D). Normally, fertilized embryos achieve a 90% epiboly stage after 10 h (Figures 6IIE). The spawned eggs of the *dnd1*$^{-/-}$ (aa) and wild-type female zebrafish showed cell death (Figures 6IIF).

To confirm the results reported above, we sampled two *dnd1*$^{-/-}$ (aa) individuals and randomly mated them with wild-type female zebrafish, and the survival rates of the spawned eggs were recorded. Additionally, we randomly selected female and male wild-type groups as the controls. In the wild-type groups, 136, 97, 122, and 178 eggs were spawned through mating, and the survival rates of the eggs averaged over 90% after 24 h (Supplementary Figure S1). In the groups of *dnd1*$^{-/-}$ (aa) (Gordon, 1997), 120, 202, 96, and 113 eggs were collected. In the groups of *dnd1*$^{-/-}$ (aa) (Ozato et al., 1992), 184, 111, 89, and 117 eggs were spawned by wild-type female zebrafish, and the eggs were unfertilized and then underwent cell death after 24 h in both the *dnd1*$^{-/-}$ (aa) (Gordon, 1997) and *dnd1*$^{-/-}$ (aa) (Ozato et al., 1992) groups (Supplementary Figure S1). This result showed that the *dnd1*-knockout zebrafish were infertile but had normal sexual characteristics.

## 3.3 Infertility control applied to fluorescent zebrafish

According to the results of the experiments above, we used wild-type zebrafish as a model to reveal the targeted mutagenesis in zebrafish *dnd1*, which may disrupt the germ cell formation, leading to a loss of germ cells and the *dnd1*$^{-/-}$ (aa) individual losing the ability to fertilize eggs. Furthermore, we applied the infertility control to fluorescent zebrafish to explore the practicalities of the infertility control technology. We analyzed the fluorescent zebrafish with the *dnd1* mutant strain by mating the cyan fluorescent zebrafish *Tg(-2.4ckmb:TcCFP13)* with the heterozygotes of pattern D (−7/WT) to generate the $F_1$ generation. Genotyping was conducted to select the heterozygotes of the individuals carrying both the heritable knockout pattern and the fluorescent pattern, inbreeding the selected individuals to generate the $F_2$ generation. Based on the results of our previous experiments, we focused on the analysis of the germ cell development and the egg fertilization ability of the mature individuals of the fluorescent *dnd1*$^{-/-}$ (aa) zebrafish. We performed histological analysis on both cyan fluorescent *dnd1*$^{+/+}$ (AA) male zebrafish and cyan fluorescent *dnd1*$^{-/-}$ (aa) zebrafish after genotyping. A completely developed testis structure was observed in the cyan fluorescent *dnd1*$^{+/+}$ (AA) candidates, in which testes developed on both sides

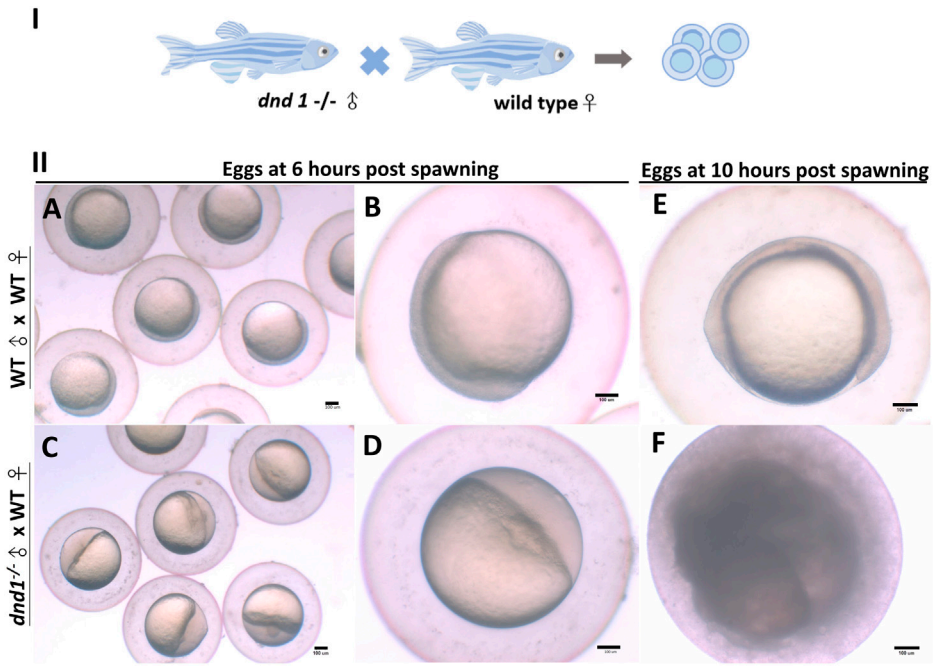



FIGURE 6
**(I)** *dnd1*$^{-/-}$ zebrafish were all male, with a normal courtship behavior. **(II)** Spawned eggs from WT female mating with *dnd1*$^{-/-}$ male and wild-type zebrafish. **(A)-(B)** Embryos spawned by wild-type zebrafish after 6 h and **(E)** after 10 h. **(C)-(D)** Eggs spawned by mutant male zebrafish and wild-type female zebrafish after 6 h and **(F)** after 10 h. Bars = 100 μm.

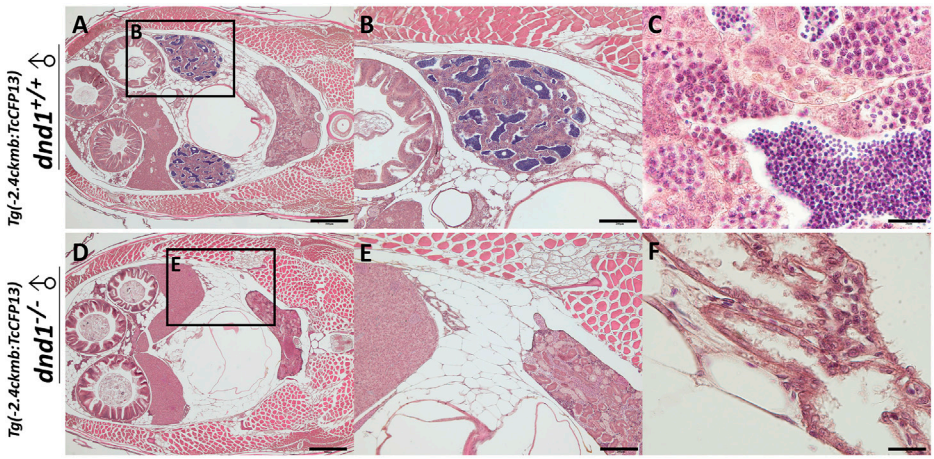
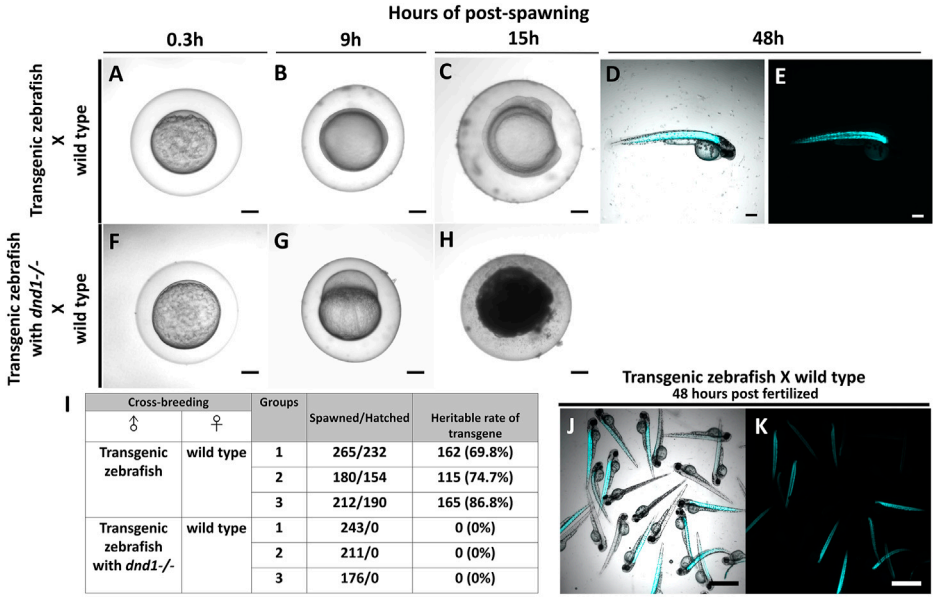



FIGURE 7
Histological analyses of fluorescent zebrafish testis structure. **(A)-(C)** *dnd1*$^{+/+}$ fluorescent zebrafish. **(D)-(F)** *dnd1*-knockout homozygote fluorescent zebrafish. Bars = 500 μm **(A, D)**, 200 μm **(B, E)**, and 20 μm **(C, F)**.

of the body (Figures 7A–C). Moreover, in the cyan fluorescent *dnd1*$^{-/-}$ (aa) zebrafish, we observed the tissue sectioning, showing no testis structure, a loss of spermatozoa, or even primary-stage germ cells in the fluorescent *dnd1*-knockout individuals (Figures 7D–F).

In case, the zebrafish which carry transgenes go into the river and mate with the wild type zebrafish, it causes widespread of the transgene as the transgenes are heritable. Therefore, we conduct the breeding experiment to simulate the escapement of the transgenic zebrafish, and to study the gene flow of the foreign gene. At the same time, we test whether the transgenes are completely inherited after performing *dnd1* mutation treatment. For easy observation, we selected the strain with muscle florescent expression to perform. Both transgenic zebrafish, transgenic zebrafish with *dnd1*$^{-/-}$ (aa), and wild type zebrafish are selected randomly for the breeding




| Cross-breeding ♂ | Cross-breeding ♀ | Groups | Spawned/Hatched | Heritable rate of transgene |
|---|---|---|---|---|
| Transgenic zebrafish | wild type | 1 | 265/232 | 162 (69.8%) |
| | | 2 | 180/154 | 115 (74.7%) |
| | | 3 | 212/190 | 165 (86.8%) |
| Transgenic zebrafish with *dnd1-/-* | wild type | 1 | 243/0 | 0 (0%) |
| | | 2 | 211/0 | 0 (0%) |
| | | 3 | 176/0 | 0 (0%) |

FIGURE 8
Transgenic zebrafish with *dnd1* mutant can effectively avoid the transgene inherited. The eggs spawned by transgenic zebrafish and wild type zebrafish **(A)** .3 h **(B)** 9 h **(C)** 15 h **(D)** 48 h of post spawning under light view **(E)** 48 h of post spawning under dark view. The eggs spawned by transgenic zebrafish with *dnd1* mutant and wild type zebrafish **(F)** .3 h **(G)** 9 h **(H)** 15 h. **(I)** The heritable rate counting of transgene in cross-breeding group. Top view of hatched offspring of cross-breeding by transgenic zebrafish and wild type zebrafish. **(J)** light view **(K)** dark view. Bars = 200 μm **(A–H)** and 500 μm **(J, K)**.

experiment. We set the breeding experiment candidate ratio to 1:1 in all groups and repeated it three times, all the eggs were collected after spawned (Figures 8A, F). We found the embryos developed normally and reached the bud stage in 9 h of post-spawning in the cross-breeding groups of transgenic zebrafish and wild type zebrafish (Figure 8B). Conversely, we observed a lack of development in the eggs that spawned by wild type females in the cross-breeding groups of transgenic zebrafish with *dnd1*$^{-/-}$ (aa) and wild type zebrafish (Figure 8G). The embryos spawned by the transgenic zebrafish and wild type zebrafish continuously developed to the somite stage after 15 h post-spawning (Figure 8C); however, the eggs that were spawned by the transgenic zebrafish with *dnd1*$^{-/-}$ (aa) and wild type zebrafish were all dead (Figure 8H). In this experiment, we found that transgenic zebrafish successfully mate with wild type zebrafish, the eggs are fertilized and develop normally during embryogenesis. Continuously, after 48 h post spawned, we found the zebrafish hatched, also some of the individuals had observed the fluorescent signal (Figures 8D, E). Conversely, in the transgenic zebrafish with the *dnd1* mutant group, the wild type female zebrafish mated with transgenic zebrafish normally. But we found the eggs were unfertilized, and show abnormal development, and undergo death within 15 h post spawning.

Further, we calculated the total number of eggs that were spawned in each cross-breeding group, and also verified the heritable rate of transgenes. In the cross-breeding groups of transgenic zebrafish and wild type zebrafish, we collected 265, 180, and 212 spawned eggs, a total of 232, 154, and 190 eggs were hatched (Figure 8I). We observed the fluorescent signal under the microscope to calculate the heritable rate of the transgene, we found the transgenic zebrafish may have 69.8%, 74.7% and 86.8% of heritage (Figures 8I–K). In the cross-breeding groups of transgenic zebrafish with *dnd1* mutant and wild type zebrafish, a total of 243, 211, and 176 eggs had collected, but all the eggs were found unfertilized (Figure 8I). From the result, the transgenic zebrafish are infertile after *dnd1* mutant, and success to avoid the transgene hereditary.

Finally, we performed a larger scale of breeding experiment to confirm the infertility effect of transgenic fish with *dnd1*$^{-/-}$ (aa), with a total of 40 transgenic fish with *dnd1*$^{-/-}$ (aa) zebrafish selected randomly (Supplementary Figure S2). We set the breeding experiment candidate ratio of transgenic cyan fluorescent zebrafish with *dnd1*$^{-/-}$ (aa) knockout and wild type female zebrafish to 1:1 (Supplementary Figure S2A). The *dnd1* homozygous mutant male fluorescent zebrafish has normal courtship behavior with wild-type female to spawning eggs (Supplementary movie). Ten transgenic fish with *dnd1*$^{-/-}$ (aa) zebrafish were examined in each experiment, and the breeding experiments were repeated four times to collect data on 40 pairs of zebrafish. Simultaneously, we used 10 pairs of wild type zebrafish as the control groups. The eggs were collected after mating, then we calculated the total number of eggs that were spawned by the wild type inbreeding groups and the groups of transgenic zebrafish with cyan fluorescent that mated with the wild type females (Supplementary Figures S2B, C, E, F). From the four breeding experiments, 1 h post-spawning in the wild-type inbreeding groups, we collected 1,347, 1,209, 1,432, and 1,156 eggs from 10 pairs of zebrafish, with a 100% survival rate (Supplementary Figure S2D); in the transgenic zebrafish with *dnd1*$^{-/-}$ (aa) zebrafish groups, we collected 1,530, 1,236, 1,282, and 1,498 eggs, with a 100% survival rate from 10 pairs of zebrafish that mated with the transgenic zebrafish with *dnd1*$^{-/-}$ (aa) zebrafish (Supplementary Figure S2G). At 6 h post-spawning, we recorded the survival status of two groups, in which no significant death was observed (Supplementary Figures S2B, E). At 24 h post-spawning, the survival rate of the four control groups was 95% or higher, with no significant number of dead eggs observed (Supplementary Figures S2C, D); in the transgenic zebrafish with

*dnd1*$^{-/-}$ (aa) zebrafish groups, we found that all of the eggs were dead (Supplementary Figures S2F, G), unlike the control. The result above shows that the infertility control technology could actually be applied to prevent the gene flow of transgene because the mutant *dnd1* individuals lost germ cells and the ability to fertilize eggs.

## 4 Discussion

As the demand for fishery resources has increased due to global population growth, marine fisheries are gradually starting to face a lack of resources. To reduce the threats relating to overfishing, to protect marine ecology, the supply of aquatic products has gradually shifted toward aquaculture (Tidwell and Allan, 2001). However, the demand for land resources and feed supplies has restrained the growth of the aquaculture industry (Cao et al., 2015). Therefore, nowadays, in the aquaculture industry, optimizing the growth, nutritional value, and disease resistance of farmed fish species, shortening their breeding times, and maximizing the production values are the main goals of development (Tidwell and Allan, 2001). To improve the quality of breeding livestock, selective breeding is commonly used in the industry. In addition, importing exotic species to adapt to cultural conditions is usually applied in the aquatic industry through global trade (Perez et al., 2003). This may increase cultivation and bring several advantages in terms of development (Gjedrem et al., 2012). However, if the artificially modified species escape from our cultural environment, the biodiversity and abundance of native species may be affected because the performance and adaptability of the selected individuals may be better than those of the native species (Lind et al., 2012).

According to previous studies, transgenic technology is a technique that could be applied in the aquatic industry for both ornamental fish and food fish, inserting exotic genes to provide the target species with the greatest performance, thus creating huge business opportunities (Gong et al., 2003; Clausen and Longo, 2012). Unfortunately, if the genetically modified species established using the transgenic technique that carry the exotic gene from the other species escape from our cultivation area, the potential risks posed to our ecosystem cannot be assessed. Thus, the application of transgenic technology has been regulated under GMO regulations. Based on the situation mentioned above, culturing sterile fish may allow the ecological risks to be avoided whenever an unexpected outflow happens. We believe that applying effective and practical infertility control technologies is important.

DND1 is a vertebrate RNA-binding protein that was first discovered in zebrafish (Weidinger et al., 2003; Kedde et al., 2007b). Previous studies have shown that the amino acid sequence of DND1 contains six functional regions, including the N-terminal region (NR), the RNA-recognition motif (RRM), and four C-terminal regions (CR 1–4) (Slanchev et al., 2009). The RRM is the most important functional region in DND1 and helps to transfer DND1 from the nucleus to the germ cell granules. While CR has ATPase activity, required for the development of PGCs and involved in the mRNA protection of downstream genes, such as *nanos3* and *tdrd7a*, mutations in this region will reduce the number of PGCs (Liu and Collodi, 2010). In previous research, the knockdown of *dnd1* expression using a morpholino in zebrafish, *Xenopus*, medaka, Atlantic cod, and other species resulted in a loss of germ cells (Weidinger et al., 2003; Youngren et al., 2005; Horvay et al., 2006; Goto et al., 2012; Northrup et al., 2012; Skugor et al., 2014; Su et al., 2014; Linhartova et al., 2015; Su et al., 2015; Hong et al., 2016). In zebrafish, the knockdown of *dnd1* resulted in a loss of germ cells due to the mis-migration and transdifferentiation of PGCs, which failed to protect the fate of the cells (Gross-Thebing et al., 2017). As *dnd1* is a critical gene that has been proven to be important in the migration and cell fate stability of PGCs, several researchers have focused on *dnd1* to perform gene disruption to explore the possibility of infertility control. In 2003, infertility control was achieved by the knockdown of the *dnd1* gene's expression using a morpholino, but performing a microinjection of one-cell fertilized eggs is an operation that is too technically complex to be applied in the industry (Weidinger et al., 2003). In 2015, researchers established the immersive bath method of Vivo-Morpholino-*dnd1* and proposed the knockdown of *dnd1*. This technique is indeed simpler to operate than microinjection, but it may be too difficult to apply in industrial applications, as the Vivo-morpholino treatment is quite expensive and non-inheritable (Wong and Zohar, 2015a). The different conditions in terms of the various fish species and egg sizes may hinder the effectiveness of the technique, and it may therefore not be able to guarantee consistency in achieving complete infertility.

In this study, gene editing was performed using CRISPR/Cas9, and the target mutagenesis site for gRNA was located before the translation of functional regions. According to the results of this study, the in-frame stop codon generated by CRISPR/Cas9 showed an abnormal translation, resulting in the *dnd1* gene showing a loss of function. This technology applied in model species, such as zebrafish, could help us to explore whether the lack of *dnd1* generated by CRISPR/Cas9 could allow complete infertility to be achieved and a complete analytical procedure to be established in a shorter period of time. We found a loss of fertilization ability in both wild-type and fluorescent zebrafish, and the targeted mutagenesis of *dnd1* using the CRISPR/Cas9 system resulted in a loss of germ cells in homozygotic individuals, which is consistent with the results presented above.

In previous studies, it was proved that *nanos3* and *tdrd7a* were protected by DND1, which competed with the binding site of the 3′ untranslated region (UTR) of the gene's mRNA to hinder its microRNA-induced degradation (Kedde et al., 2007a; Aguero et al., 2017). We analyzed *dnd1*-related genes, such as *nanos3*, *tdrd7a*, and *ddx4*, in *dnd1*-knockout individuals. In *dnd1*-knockout homozygotic individuals, germ-cell-related genes resulted in a lack of expression, but there was a gonadal somatic cell expression of the *amh* gene. We conjecture that the knockout of *dnd1* caused a loss of germ cells, but with no significant effect on the gonadal somatic cells. According to research on *dnd1* knockdown in medaka, injecting different concentrations of a morpholino leads to a direct effect on the number of PGCs (Hong et al., 2016). However, we did not observe germ-cell-related gene downregulation or histological evidence in heterozygotic individuals. We could only determine whether there were differences in the gene expression levels and the fertility capacity to determine whether there was a significant difference between the heterozygotes and wild types.

In this study, we found that the all-male phenotype in *dnd1*-knockout zebrafish by CRISPR/Cas9 was complete sterile, although the males still maintained normal courtship behavior. In the study of *dnd1*-knockdown zebrafish, we observed the same all-male phenomenon (Pradhan and Olsson, 2018). According to previous research, a lack of germ cells in zebrafish, medaka, and tilapia leads to their sex determination in the all-male phenotype, as the number of germ cells may affect the differentiation

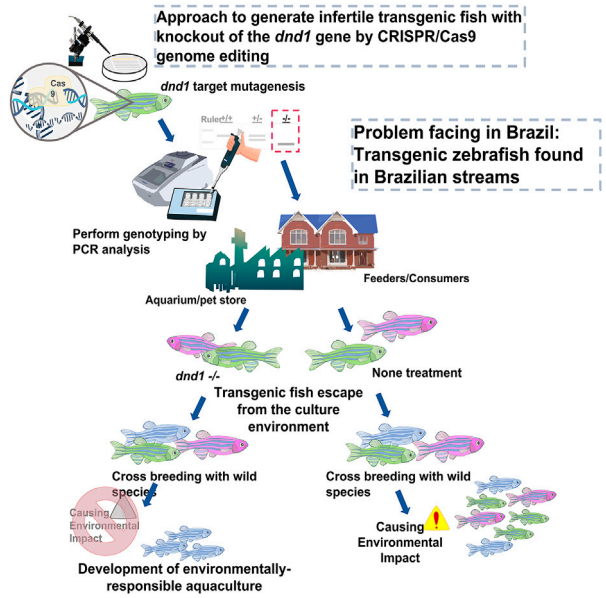



FIGURE 9
Approach to generate infertile transgenic fluorescent zebrafish with knockout out of the *dnd1* gene by CRISPR/Cas9 genome editing. We demonstrate that the *dnd1* knockout by CRISPR/Cas9 genome editing could be applied to the infertility control of transgenic fluorescent zebrafish to prevent them from cross-breeding with wild species when they escape into the wild field. It can promote the development of an environmentally responsible aquaculture.

mechanism (Slanchev et al., 2005; Kurokawa et al., 2007; Li et al., 2014). In our study, CRISPR/Cas9 technology, which needs only prepare a single guide RNA and Cas9 mRNA to induce dsDNA break for genome editing, was used to knockout *dnd1* gene for complete infertility control of transgenic fish. The CRISPR/Cas9 genome editing technology is a more easy, cheap and precise way than ZFNs and TALENs. As zebrafish as a model species, to establish an integrated infertility control procedure by CRISPR/Cas9 genome editing from the development to the analysis further is necessary. Also, this study mainly focused on the application of infertility control on transgenic fish, and we provided evidence to emphasize the practicality of this technique. We established *dnd1*-knockout zebrafish using CRISPR/Cas9 gene editing technology and achieved complete infertility in homozygotic individuals of transgenic cyan fluorescent zebrafish. While this application of preparatory work to perform gene editing is complicated, as long as heterozygous mutants are established, and the stable heritable knockout genome sequences are selected, followed by the interbreeding of heterozygotes, homozygotic individuals can be analyzed by fin cleaving, PCR amplification and genotyping (Figure 9). Besides the generation of completely infertile fish, infertility species may be used as germ cell transplantation carriers to produce the germ cells of different species (Saito et al., 2008; Yazawa et al., 2010; Yazawa et al., 2013; Pacchiarini et al., 2014; Morita et al., 2015; Lujic et al., 2018). However, the idea of infertility control by knockout strategy seems to be impractical when it is put into the application of in aquaculture scale. As the verification of genotype may require PCR amplification procedure and the inheritable homozygote ratio which is around 1/4 of the total offspring. As fertile *dnd1* heterozygous mutants still can breed, they have to be kept in breeding company under strict regulation to prevent from escape. Only sterile transgenic fluorescent zebrafish with *dnd1* homozygous mutations can be sold to local aquariums and consumers. There is no practical application if fertility cannot be restored in *dnd1* homozygous mutants for mass production of sterile offspring. In Atlantic salmon, rescue of germ cells in *dnd1* knockout embryos by CRISPR/Cas9 opened possibility to produce inherited sterility (Guralp et al., 2020). Development of practical mass production technology of sterile fish still need to be conquered by more studies and efforts. Nevertheless, practicality and stability are the most important elements in the infertility control technique. We propose that this technique may be suitable for the selective breeding of fish species to achieve a higher product value or to achieve transgenic individuals, which infertility control must be applied to ensure that the transgenic individuals are completely sterile.

Nowadays, the development of biotechnology is mature, and many important breeding fish species have developed breeding technologies with great application potential. In addition to being applied to the industry, these technologies must also pass numerous regulatory and analytical tests. Especially for the artificial selective breeding or genetic modified fish, the most concern of us is to protect our natural primitive ecology by avoiding the foreign gene or dominant species from invading or affecting the wild population, whereas the application of infertility control technique is one of the keys. In this study, we introduce a complete procedure from the development of *dnd1* knockout candidates to the analyses of stable heritable rates of homozygotes zebrafish, with multiple breeding experiments to confirm the sterile fish which did not maintain their reproductive ability. It is important to develop and employ effective infertility control in industry practice, as this may allow us to protect our ecosystem and build up a responsible aquaculture environment. Although our study achieved infertility control by genome editing of the *dead end* gene, which has been reported in zebrafish by ZFNs, we still believed that using new gene editing technology 'CRISPR/Cas9′ may help the progress of infertility control applicated in the industry, as the *dead end* gene is conserved between vertebrates and the accomplishment of the target mutagenesis is more simple, faster and precise. Zebrafish is the most important model species, and the completeness and systematic process description are vitally important for the reference of other species.

## 5 Conclusion

To avoid the ecological risks associated with escaping transgenic fish into the aquatic environment, the development of an effective sterilization strategy is essential. *Dead end* (*dnd1*) is a critical gene that plays an essential role in the migration, survival, and cell fate maintenance of primordial germ cells (PGCs). In this study, we report the targeted mutagenesis of the *dnd1* gene in zebrafish by using CRISPR/Cas9 genome editing technology to achieve complete sterility in *dnd1*-gene-knockout non-transgenic zebrafish and transgenic fluorescent zebrafish, as the model for the infertility control in farmed ornamental fish and transgenic fish, especially in transgenic fluorescent ornamental fish.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Ethics statement

The animal study was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of National Taiwan Ocean University under Approval No. 108042.

## Author contributions

Conceptualization, H-YG and J-LW; Methodology, W-KC; Validation, W-KC; Formal analysis, W-KC; Investigation, W-KC and S-CH; Resources, H-YG and C-FC; Data curation, W-KC and S-CH; Writing—original draft preparation, W-KC; Writing—review and editing, H-YG and C-FC; Visualization, J-LW and C-FC; Supervision, H-YG and C-FC; Project administration, H-YG; Funding acquisition, H-YG. All authors have read and agreed to the published version of the manuscript.

## Funding

The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

## Conflict of interest

This study received funding (grant No.:MOST 109-2622-B-019-002-CC1) from the Ministry of Science and Technology, Taiwan, and Jy Lin Trading Co., Ltd., in Taiwan, to H-YG.

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

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1029200/full#supplementary-material

## References

Aguero, T., Jin, Z., Chorghade, S., Kalsotra, A., King, M. L., and Yang, J. (2017). Maternal Dead-end 1 promotes translation of nanos1 by binding the eIF3 complex. *Development* 144 (20), 3755–3765. doi:10.1242/dev.152611

Arai, K. (2001). Genetic improvement of aquaculture finfish species by chromosome manipulation techniques in Japan. *Aquaculture* 197 (1-4), 205–228. doi:10.1016/s0044-8486(01)00588-9

Aramaki, S., Sato, F., Kato, T., Soh, T., Kato, Y., and Hattori, M. A. (2007). Molecular cloning and expression of dead end homologue in chicken primordial germ cells. *Cell Tissue Res.* 330 (1), 45–52. doi:10.1007/s00441-007-0435-1

Auer, T. O., Duroure, K., De Cian, A., Concordet, J. P., and Del Bene, F. (2014). Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* 24 (1), 142–153. doi:10.1101/gr.161638.113

Baloch, A. R., Franek, R., Saito, T., and Psenicka, M. (2021). Dead-end (dnd) protein in fish-a review. *Fish. Physiol. Biochem.* 47 (3), 777–784. doi:10.1007/s10695-018-0606-x

Baloch, A. R., Franěk, R., Tichopád, T., Fučíková, M., Rodina, M., and Pšenička, M. (2019). *Dnd1* knockout in sturgeons by CRISPR/Cas9 generates germ cell free host for surrogate production. *Anim. (Basel)* 179 (4), 174. doi:10.3390/ani9040174

Benfey, T. J., Dye, H. M., SolarII, and Donaldson, E. M. (1989). The growth and reproductive endocrinology of adult triploid Pacific salmonids. *Fish. Physiol. Biochem.* 6 (2), 113–120. doi:10.1007/BF01875483

Bhattacharya, D., and Van Meir, E. G. (2019). A simple genotyping method to detect small CRISPR-Cas9 induced indels by agarose gel electrophoresis. *Sci. Rep.* 9 (1), 4437. doi:10.1038/s41598-019-39950-4

Blazer, V. S. (2002). Histopathological assessment of gonadal tissue in wild fishes. *Fish Physiology Biochem.* 26 (1), 85–101. doi:10.1023/a:1023332216713

Braat, A. K., Zandbergen, T., Van de Water, S., Goos, H. J. T., and Zivkovic, D. (1999). Characterization of zebrafish primordial germ cells: Morphology and early distribution of vasa RNA. *Dev. Dynam* 216 (2), 153–167. doi:10.1002/(SICI)1097-0177(199910)216:2<153::AID-DVDY6>3.0.CO;2-1

Brend, T., and Holley, S. A. (2009). Zebrafish whole mount high-resolution double fluorescent <em>*in situ*</em> hybridization. *J. Vis. Exp.* 25, 1229. doi:10.3791/1229

Cal, R. M., idal, S., Gomez, C., Alvarez-Blazquez, B., Martinez, P., and Piferrer, F. (2006). Growth and gonadal development in diploid and triploid turbot (*Scophthalmus maximus*). *Aquaculture* 251 (1), 99–108. doi:10.1016/j.aquaculture.2005.05.010

Cao, L., Naylor, R., Henriksson, P., Leadbitter, D., Metian, M., Troell, M., et al. (2015). Global food supply. China's aquaculture and the world's wild fisheries. *Science* 347 (6218), 133–135. doi:10.1126/science.1260149

Castrillon, D. H., Quade, B. J., Wang, T. Y., Quigley, C., and Crum, C. P. (2000). The human VASA gene is specifically expressed in the germ cell lineage. *P Natl. Acad. Sci. U. S. A.* 97 (17), 9585–9590. doi:10.1073/pnas.160274797

Chapman, J. R., Taylor, M. R. G., and Boulton, S. J. (2012). Playing the end game: DNA double-strand break repair pathway choice. *Mol. Cell* 47 (4), 497–510. doi:10.1016/j.molcel.2012.07.029

Chudakov, D. M., Matz, M. V., Lukyanov, S., and Lukyanov, K. A. (2010). Fluorescent proteins and their applications in imaging living cells and tissues. *Physiol. Rev.* 90 (3), 1103–1163. doi:10.1152/physrev.00038.2009

Clausen, R., and Longo, S. B. (2012). The tragedy of the commodity and the farce of AquAdvantage Salmon. *Dev. Change* 43 (1), 229–251. doi:10.1111/j.1467-7660.2011.01747.x

Cook, J. T., McNiven, M. A., Richardson, G. F., and Sutterlin, A. M. (2000). Growth rate, body composition and feed digestibility/conversion of growth-enhanced transgenic Atlantic salmon (*Salmo salar*). *Aquaculture* 188 (1-2), 15–32. doi:10.1016/s0044-8486(00)00331-8

Doudna, J. A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346, 1258096. doi:10.1126/science.1258096

Du, S. J., Gong, Z. Y., Fletcher, G. L., Shears, M. A., King, M. J., Idler, D. R., et al. (1992). Growth enhancement in transgenic Atlantic salmon by the use of an "all fish" chimeric growth hormone gene construct. *Biotechnol. (N Y)* 10 (2), 176–181. doi:10.1038/nbt0292-176

Extavour, C. G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: Epigenesis and preformation. *Development* 130 (24), 5869–5884. doi:10.1242/dev.00804

Fujihara, R., Katayama, N., Sadaie, S., Miwa, M., Sanchez Matias, G. A., Ichida, K., et al. (2022). Production of germ cell-less rainbow trout by dead end gene knockout and their use as recipients for germ cell transplantation. *Mar. Biotechnol. (NY)* 24 (2), 417–429. doi:10.1007/s10126-022-10128-w

Garneau, J. E., Dupuis, M. E., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., et al. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468 (7320), 67–71. doi:10.1038/nature09523

Gjedrem, T., Robinson, N., and Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* 350, 117–129. doi:10.1016/j.aquaculture.2012.04.008

Gong, H. Y., Hu, S. Y., and Wu, J. L. (2015). Progress of transgenic fluorescent ornamental fish and development of its infertility control technologies. *Agric. Biotechnol. Ind. Q.* 43, 34–48.

Gong, Z., Ju, B., and Wan, H. (2001). Green fluorescent protein (GFP) transgenic fish and their applications. *Genetica* 111 (1-3), 213–225. doi:10.1023/a:1013796810782

Gong, Z., Wan, H., Tay, T. L., Wang, H., Chen, M., and Yan, T. (2003). Development of transgenic fish for ornamental and bioreactor by strong expression of fluorescent proteins in the skeletal muscle. *Biochem. Biophys. Res. Commun.* 308 (1), 58–63. doi:10.1016/s0006-291x(03)01282-8

Gordon, J. W. (1997). Transgenic technology and laboratory animal science. *ILAR J.* 38 (1), 32–41. doi:10.1093/ilar.38.1.32

Goto, R., Saito, T., Takeda, T., Fujimoto, T., Takagi, M., Arai, K., et al. (2012). Germ cells are not the primary factor for sexual fate determination in goldfish. *Dev. Biol.* 370 (1), 98–109. doi:10.1016/j.ydbio.2012.07.010

Gross-Thebing, T., Yigit, S., Pfeiffer, J., Reichman-Fried, M., Bandemer, J., Ruckert, C., et al. (2017). The vertebrate protein dead end maintains primordial germ cell fate by inhibiting somatic differentiation. *Dev. Cell* 43 (6), 704–715. doi:10.1016/j.devcel.2017.11.019

Guralp, H., Skaftnesmo, K. O., Kjaerner-Semb, E., Straume, A. H., Kleppe, L., Schulz, R. W., et al. (2020). Rescue of germ cells in dnd crispant embryos opens the possibility to produce inherited sterility in Atlantic salmon. *Sci. Rep.* 10 (1), 18042. doi:10.1038/s41598-020-74876-2

Hartung, O., Forbes, M. M., and Marlow, F. L. (2014). Zebrafish vasa is required for germ-cell differentiation and maintenance. *Mol. Reproduction Dev.* 81 (10), 946–961. doi:10.1002/mrd.22414

Her, G. M., Chiang, C. C., Chen, W. Y., and Wu, J. L. (2003). *In vivo* studies of liver-type fatty acid binding protein (L-FABP) gene expression in liver of transgenic zebrafish (*Danio rerio*). *FEBS Lett.* 538 (1-3), 125–133. doi:10.1016/s0014-5793(03)00157-1

Hong, N., Li, M. Y., Yuan, Y. M., Wang, T. S., Yi, M. S., Xu, H. Y., et al. (2016). Dnd is a critical specifier of primordial germ cells in the medaka fish. *Stem Cell Rep.* 6 (3), 411–421. doi:10.1016/j.stemcr.2016.01.002

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327 (5962), 167–170. doi:10.1126/science.1179555

Horvay, K., Claussen, M., Katzer, M., Landgrebe, J., and Pieler, T. (2006). Xenopus Dead end mRNA is a localized maternal determinant that serves a conserved function in germ cell development. *Dev. Biol.* 291 (1), 1–11. doi:10.1016/j.ydbio.2005.06.013

Hu, W., Wang, Y., and Zhu, Z. (2007). Progress in the evaluation of transgenic fish for possible ecological risk and its containment strategies. *Sci. China C Life Sci.* 50 (5), 573–579. doi:10.1007/s11427-007-0089-y

Hwang, W. Y., Fu, Y. F., Reyon, D., Maeder, M. L., Tsai, S. Q., Sander, J. D., et al. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* 31 (3), 227–229. doi:10.1038/nbt.2501

Jao, L. E., Wente, S. R., and Chen, W. B. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *P Natl. Acad. Sci. U. S. A.* 110 (34), 13904–13909. doi:10.1073/pnas.1308335110

Ju, B. S., Xu, X. F., He, J. Y., Liao, J., Yan, T., Hew, C. L., et al. (1999). Faithful expression of green fluorescent protein (GFP) in transgenic zebrafish embryos under control of zebrafish gene promoters. *Dev. Genet.* 25 (2), 158–167. doi:10.1002/(SICI)1520-6408(1999)25:2<158::AID-DVG10>3.0.CO;2-6

Kawakami, K. (2007). Tol2: A versatile gene transfer vector in vertebrates. *Genome Biol.* 8 (1), S7. doi:10.1186/gb-2007-8-s1-s7

Kedde, M., Strasser, M. J., Boldajipour, B., Oude Vrielink, J. A., Slanchev, K., le Sage, C., et al. (2007). RNA-binding protein *Dnd1* inhibits microRNA access to target mRNA. *Cell* 131 (7), 1273–1286. doi:10.1016/j.cell.2007.11.034

Kedde, M., Strasser, M. J., Boldajipour, B., Vrielink, J. A. F. O., Le Sage, C., Nagel, R., et al. (2007). RNA-binding protein *Dnd1* inhibits microRNA access to target mRNA. *Cell* 131 (7), 1273–1286. doi:10.1016/j.cell.2007.11.034

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874. doi:10.1093/molbev/msw054

Kurokawa, H., Saito, D., Nakamura, S., Katoh-Fukui, Y., Ohta, K., Baba, T., et al. (2007). Germ cells are essential for sexual dimorphism in the medaka gonad. *Proc. Natl. Acad. Sci. U. S. A.* 104 (43), 16958–16963. doi:10.1073/pnas.0609932104

Labun, K., Montague, T. G., Gagnon, J. A., Thyme, S. B., and Valen, E. (2016). CHOPCHOP v2: A web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* 44 (W1), W272–W276. doi:10.1093/nar/gkw398

Li, M., Yang, H., Zhao, J., Fang, L., Shi, H., Li, M., et al. (2014). Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. *Genetics* 197 (2), 591–599. doi:10.1534/genetics.114.163667

Li, M. Y., Hong, N., Xu, H. Y., Yi, M. S., Li, C. M., Gui, J. F., et al. (2009). Medaka vasa is required for migration but not survival of primordial germ cells. *Mech. Dev.* 126 (5-6), 366–381. doi:10.1016/j.mod.2009.02.004

Li, Q., Fujii, W., Naito, K., and Yoshizaki, G. (2017). Application of dead end-knockout zebrafish as recipients of germ cell transplantation. *Mol. Reproduction Dev.* 84 (10), 1100–1111. doi:10.1002/mrd.22870

Lind, C. E., Ponzoni, R. W., Nguyen, N. H., and Khaw, H. L. (2012). Selective breeding in fish and conservation of genetic resources for aquaculture. *Reprod. Domest. Anim.* 47 (4), 255–263. doi:10.1111/j.1439-0531.2012.02084.x

Linhartova, Z., Saito, T., Kaspar, V., Rodina, M., Praskova, E., Hagihara, S., et al. (2015). Sterilization of sterlet *Acipenser ruthenus* by using knockdown agent, antisense morpholino oligonucleotide, against dead end gene. *Theriogenology* 84 (7), 1246–1255. doi:10.1016/j.theriogenology.2015.07.003

Liu, L. X., Hong, N., Xu, H. Y., Li, M. Y., Yan, Y., Purwanti, Y., et al. (2009). Medaka dead end encodes a cytoplasmic protein and identifies embryonic and adult germ cells. *Gene Expr. Patterns* 9 (7), 541–548. doi:10.1016/j.gep.2009.06.008

Liu, W., and Collodi, P. (2010). Zebrafish dead end possesses ATPase activity that is required for primordial germ cell development. *FASEB J.* 24 (8), 2641–2650. doi:10.1096/fj.09-148403

Lujic, J., Marinovic, Z., Bajec, S. S., Djurdjevic, I., Urbanyi, B., and Horvath, A. (2018). Interspecific germ cell transplantation: A new light in the conservation of valuable balkan trout genetic resources? *Fish. Pysiol Biochem.* 44 (6), 1487–1498. doi:10.1007/s10695-018-0510-4

Maclean, N. (1993). Transgenic induction in Salmonid and Tilapia fish. *Methods Mol. Biol.* 18, 95–107. doi:10.1385/0-89603-245-0:95

Magalhaes, A. L. B., Brito, M. F. G., and Silva, L. G. M. (2022). The fluorescent introduction has begun in the southern hemisphere: Presence and life-history strategies of the transgenic zebrafish *Danio rerio* (cypriniformes: Danionidae) in Brazil. *Stud. Neotrop Fauna E.* doi:10.1080/01650521.2021.2024054

Meeker, N. D., Hutchinson, S. A., Ho, L., and Trecle, N. S. (2007). Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *Biotechniques* 43 (5), 610, 612, 614–+. doi:10.2144/000112619

Miki, M., Ohishi, N., Nakamura, E., Furumi, A., and Mizuhashi, F. (2018). Improved fixation of the whole bodies of fish by a double-fixation method with formalin solution and Bouin's fluid or Davidson's fluid. *J. Toxicol. Pathol.* 31 (3), 201–206. doi:10.1293/tox.2018-0001

Morita, T., Morishima, K., Miwa, M., Kumakura, N., Kudo, S., Ichida, K., et al. (2015). Functional sperm of the yellowtail (*Seriola quinqueradiata*) were produced in the small-bodied surrogate, Jack Mackerel (Trachurus japonicus). *Mar. Biotechnol.* 17 (5), 644–654. doi:10.1007/s10126-015-9657-5

Moutinho, S. (2022). Transgenic fish invades Brazilian streams. *Science* 375 (6582), 704–705. doi:10.1126/science.ada1331

Muir, W. M., and Howard, R. D. (1999). Possible ecological risks of transgenic organism release when transgenes affect mating success: Sexual selection and the Trojan gene hypothesis. *P Natl. Acad. Sci. U. S. A.* 96 (24), 13853–13856. doi:10.1073/pnas.96.24.13853

Muir, W. M. (2004). The threats and benefits of GM fish. *EMBO Rep.* 5 (7), 654–659. doi:10.1038/sj.embor.7400197

Nagasawa, K., Fernandes, J. M. O., Yoshizaki, G., Miwa, M., and Babiak, I. (2013). Identification and migration of primordial germ cells in Atlantic salmon, *Salmo salar*: Characterization of Vasa, Dead End, and Lymphocyte antigen 75 genes. *Mol. Reproduction Dev.* 80 (2), 118–131. doi:10.1002/mrd.22142

Northrup, E., Zschemisch, N. H., Eisenblatter, R., Glage, S., Wedekind, D., Cuppen, E., et al. (2012). The ter mutation in the rat *Dnd1* gene initiates gonadal teratomas and infertility in both genders. *Plos One* 7 (5), e38001. doi:10.1371/journal.pone.0038001

Ozato, K., Wakamatsu, Y., and Inoue, K. (1992). Medaka as a model of transgenic fish. *Mol. Mar. Biol. Biotechnol.* 1 (4-5), 346–354.

Pacchiarini, T., Sarasquete, C., and Cabrita, E. (2014). Development of interspecies testicular germ-cell transplantation in flatfish. *Reprod. Fertil. Dev.* 26 (5), 690–702. doi:10.1071/RD13103

Perez, J. E., Alfonsi, C., Nirchio, M., Munoz, C., and Gomez, J. A. (2003). The introduction of exotic species in aquaculture: A solution or part of the problem? *Interciencia* 28 (4), 234–238.

Piferrer, F., Beaumont, A., Falguiere, J. C., Flajshans, M., Haffray, P., and Colombo, L. (2009). Polyploid fish and shellfish: Production, biology and applications to aquaculture for performance improvement and genetic containment. *Aquaculture* 293 (3-4), 125–156. doi:10.1016/j.aquaculture.2009.04.036

Pradhan, A., and Olsson, P. E. (2018). Germ cell depletion in zebrafish leads to incomplete masculinization of the brain. *Gen. Comp. Endocrinol.* 265, 15–21. doi:10.1016/j.ygcen.2018.02.001

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8 (11), 2281–2308. doi:10.1038/nprot.2013.143

Raz, E. (2003). Primordial germ-cell development: The zebrafish perspective. *Nat. Rev. Genet.* 4 (9), 690–700. doi:10.1038/nrg1154

Raz, E. (2000). The function and regulation of vasa-like genes in germ-cell development. *Genome Biol.* 1 (3), REVIEWS1017. doi:10.1186/gb-2000-1-3-reviews1017

Saito, T., Goto-Kazeto, R., Arai, K., and Yamaha, E. (2008). Xenogenesis in teleost fish through generation of germ-line chimeras by single primordial germ cell transplantation. *Biol. Reproduction* 78 (1), 159–166. doi:10.1095/biolreprod.107.060038

Sawamura, R., Osafune, N., Murakami, T., Furukawa, F., and Kitano, T. (2017). Generation of biallelic F0 mutants in medaka using the CRISPR/Cas9 system. *Genes cells.* 22 (8), 756–763. doi:10.1111/gtc.12511

Skugor, A., Tveiten, H., Krasnov, A., and Andersen, O. (2014). Knockdown of the germ cell factor Dead end induces multiple transcriptional changes in Atlantic cod (*Gadus morhua*) hatchlings. *Anim. Reprod. Sci.* 144 (3-4), 129–137. doi:10.1016/j.anireprosci.2013.12.010

Slanchev, K., Stebler, J., de la Cueva-Mendez, G., and Raz, E. (2005). Development without germ cells: The role of the germ line in zebrafish sex differentiation. *P Natl. Acad. Sci. U. S. A.* 102 (11), 4074–4079. doi:10.1073/pnas.0407475102

Slanchev, K., Stebler, J., Goudarzi, M., Cojocaru, V., Weidinger, G., and Raz, E. (2009). Control of Dead end localization and activity--implications for the function of the protein in antagonizing miRNA function. *Mech. Dev.* 126 (3-4), 270–277. doi:10.1016/j.mod.2008.10.006

Stewart, C. N., Jr. (2006). Go with the glow: Fluorescent proteins to light transgenic organisms. *Trends Biotechnol.* 24 (4), 155–162. doi:10.1016/j.tibtech.2006.02.002

Stigebrandt, A., Aure, J., Ervik, A., and Hansen, P. K. (2004). Regulating the local environmental impact of intensive marine fish farming. *Aquaculture* 234 (1-4), 239–261. doi:10.1016/j.aquaculture.2003.11.029

Su, B. F., Peatman, E., Shang, M., Thresher, R., Grewe, P., Patil, J., et al. (2014). Expression and knockdown of primordial germ cell genes, vasa, nanos and dead end in common carp (*Cyprinus carpio*) embryos for transgenic sterilization and reduced sexual maturity. *Aquaculture* 420, S72–S84. doi:10.1016/j.aquaculture.2013.07.008

Su, B. F., Shang, M., Grewe, P. M., Patil, J. G., Peatman, E., Perera, D. A., et al. (2015). Suppression and restoration of primordial germ cell marker gene expression in channel catfish, *Ictalurus punctatus*, using knockdown constructs regulated by copper transport protein gene promoters: Potential for reversible transgenic sterilization. *Theriogenology* 84 (9), 1499–1512. doi:10.1016/j.theriogenology.2015.07.037

Suster, M. L., Kikuta, H., Urasaki, A., Asakawa, K., and Kawakami, K. (2009). Transgenesis in zebrafish with the tol2 transposon system. *Methods Mol. Biol.* 561, 41–63. doi:10.1007/978-1-60327-019-9_3

Tidwell, J. H., and Allan, G. L. (2001). Fish as food: aquaculture's contribution - ecological and economic impacts and contributions of fish farming and capture fisheries. *Embo Rep.* 2 (11), 958–963. doi:10.1093/embo-reports/kve236

Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., et al. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* 435 (7042), 646–651. doi:10.1038/nature03556

Wagner, E. J., Arndt, R. E., Routledge, M. D., Latremouille, D., and Mellenthin, R. F. (2006). Comparison of hatchery performance, agonistic behavior, and poststocking survival between diploid and triploid rainbow trout of three different Utah strains. *N. Am. J. Aquacult* 68 (1), 63–73. doi:10.1577/a05-026.1

Wang, T. (2015). *Gene editing of* dnd *by TALENs in medaka embryos and its role in PGC formation*. Singapore: Ph.D. thesis. Department of Biological Sciencs, National University of Singapore.

Wang, T., and Hong, Y. (2014). Direct gene disruption by TALENs in medaka embryos. *Gene* 543 (1), 28–33. doi:10.1016/j.gene.2014.04.013

Wargelius, A., Leininger, S., Skaftnesmo, K. O., Kleppe, L., Andersson, E., Taranger, G. L., et al. (2016). Dnd knockout ablates germ cells and demonstrates germ cell independent sex differentiation in Atlantic salmon. *Sci. Rep.* 18, 21284. doi:10.1038/srep21284

Weidinger, G., Stebler, J., Slanchev, K., Dumstrei, K., Wise, C., Lovell-Badge, R., et al. (2003). Dead end, a novel vertebrate germ plasm component, is required for zebrafish primordial germ cell migration and survival. *Curr. Biol.* 13 (16), 1429–1434. doi:10.1016/s0960-9822(03)00537-2

Weidinger, G., Wolke, U., Koprunner, M., Klinger, M., and Raz, E. (1999). Identification of tissues and patterning events required for distinct steps in early migration of zebrafish primordial germ cells. *Development* 126 (23), 5295–5307. doi:10.1242/dev.126.23.5295

Wong, A. C., and Van Eenennaam, A. L. (2008). Transgenic approaches for the reproductive containment of genetically engineered fish. *Aquaculture* 275 (1-4), 1–12. doi:10.1016/j.aquaculture.2007.12.026

Wong, T. T., and Collodi, P. (2013). Inducible sterilization of zebrafish by disruption of primordial germ cell migration. *Plos One* 8 (6), e68455. doi:10.1371/journal.pone.0068455

Wong, T. T., and Zohar, Y. (2015). Production of reproductively sterile fish by a non-transgenic gene silencing technology. *Sci. Rep-Uk* 5, 15822. doi:10.1038/srep15822

Wong, T. T., and Zohar, Y. (2015). Production of reproductively sterile fish: A mini-review of germ cell elimination technologies. *Gen. Comp. Endocr.* 221, 3–8. doi:10.1016/j.ygcen.2014.12.012

Xu, H. Y., Li, M. Y., Gui, J. F., and Hong, Y. H. (2010). Fish germ cells. *Sci. China Life Sci.* 53 (4), 435–446. doi:10.1007/s11427-010-0058-8

Yamaha, E., Goto-Kazeto, R., Saito, T., Kawakami, Y., Fujimoto, T., Adachi, S., et al. (2010). Primordial germ cell in teleost fish with special references to its specification and migration. *J. Appl. Ichthyol.* 26 (5), 816–822. doi:10.1111/j.1439-0426.2010.01548.x

Yazawa, R., Takeuchi, Y., Higuchi, K., Yatabe, T., Kabeya, N., and Yoshizaki, G. (2010). Chub mackerel gonads support colonization, survival, and proliferation of intraperitoneally transplanted xenogenic germ cells. *Biol. Reprod.* 82 (5), 896–904. doi:10.1095/biolreprod.109.081281

Yazawa, R., Takeuchi, Y., Morita, T., Ishida, M., and Yoshizaki, G. (2013). The Pacific bluefin tuna (Thunnus orientalis) dead end gene is suitable as a specific molecular marker of type A spermatogonia. *Mol. Reprod. Dev.* 80 (10), 871–880. doi:10.1002/mrd.22224

Yoon, C., Kawakami, K., and Hopkins, N. (1997). Zebrafish vasa homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells. *Development* 124 (16), 3157–3165. doi:10.1242/dev.124.16.3157

Yoshizaki, G., Sakatani, S., Tominaga, H., and Takeuchi, T. (2000). Cloning and characterization of a vasa-like gene in rainbow trout and its expression in the germ cell lineage. *Mol. Reprod. Dev.* 55 (4), 364–371. doi:10.1002/(SICI)1098-2795(200004)55:4<364::AID-MRD2>3.0.CO;2-8

Youngren, K. K., Coveney, D., Peng, X., Bhattacharya, C., Schmidt, L. S., Nickerson, M. L., et al. (2005). The Ter mutation in the dead end gene causes germ cell loss and testicular germ cell tumours. *Nature* 435 (7040), 360–364. doi:10.1038/nature03595

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 19 January 2023
DOI 10.3389/fgene.2023.1097825



OPEN ACCESS

EDITED BY
Yuzine Esa,
Putra Malaysia University, Malaysia

REVIEWED BY
Qian Wang,
Yellow Sea Fisheries Research Institute (CAFS), China
Mingyou Li,
Shanghai Ocean University, China
Min Tao,
Hunan Normal University, China
Wei Hu,
State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology (CAS), China

*CORRESPONDENCE
Zhi-Hui Sun,
sunzhihui@dlou.edu.cn

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

RECEIVED 14 November 2022
ACCEPTED 06 January 2023
PUBLISHED 19 January 2023

CITATION
Liu B-Z, Cong J-J, Su W-Y, Hao Z-L, Sun Z-H and Chang Y-Q (2023), Identification and functional analysis of *Dmrt1* gene and the *SoxE* gene in the sexual development of sea cucumber, *Apostichopus japonicus*. *Front. Genet.* 14:1097825. doi: 10.3389/fgene.2023.1097825



# Identification and functional analysis of *Dmrt1* gene and the *SoxE* gene in the sexual development of sea cucumber, *Apostichopus japonicus*

Bing-Zheng Liu, Jing-Jing Cong, Wei-Yi Su, Zhen-Lin Hao, Zhi-Hui Sun* and Ya-Qing Chang

Key Laboratory of Mariculture and Stock Enhancement in North China's Sea, Ministry of Agriculture and Rural Affairs, Dalian Ocean University, Dalian, Liaoning, China

Members of the Doublesex and Mab-3-related transcription factor (*Dmrt*) gene family handle various vital functions in several biological processes, including sex determination/differentiation and gonad development. *Dmrt1* and *Sox9* (*SoxE* in invertebrates) exhibit a very conserved interaction function during testis formation in vertebrates. However, the dynamic expression pattern and functional roles of the *Dmrt* gene family and *SoxE* have not yet been identified in any echinoderm species. Herein, five members of the *Dmrt* gene family (*Dmrt1*, *2*, *3a*, *3b* and *5*) and the ancestor *SoxE* gene were identified from the genome of *Apostichopus japonicus*. Expression studies of *Dmrt* family genes and *SoxE* in different tissues of adult males and females revealed different expression patterns of each gene. Transcription of *Dmrt2*, *Dmrt3a* and *Dmrt3b* was higher expressed in the tube feet and coelomocytes instead of in gonadal tissues. The expression of *Dmrt1* was found to be sustained throughout spermatogenesis. Knocking-down of *Dmrt1* by means of RNA interference (RNAi) led to the downregulation of *SoxE* and upregulation of the ovarian regulator *foxl2* in the testes. This indicates that *Dmrt1* may be a positive regulator of *SoxE* and may play a role in the development of the testes in the sea cucumber. The expression level of *SoxE* was higher in the ovaries than in the testes, and knocking down of *SoxE* by RNAi reduced *SoxE* and *Dmrt1* expression but conversely increased the expression of *foxl2* in the testes. In summary, this study indicates that *Dmrt1* and *SoxE* are indispensable for testicular differentiation, and *SoxE* might play a functional role during ovary differentiation in the sea cucumber.

KEYWORDS

**RNAi, sea cucumber, gonad development, DMRT, SoxE**

## 1 Introduction

The sea cucumber *A. japonicus* is an important nutritional seafood and is regarded as a precious traditional Chinese medicine (Yan J. et al., 2013; Yu Z. et al., 2014; Yang et al., 2015). The sea cucumber is widely distributed in the coastal regions of China, Far Eastern Russia, Japan, and Korea (Tian et al., 2017; Chen and Chang, 2015; Yan J. et al., 2013; Yu Z. et al., 2014), and its scale of breeding has continued to expand. Compared with the female sea cucumber, the male might have an advantage in terms of immunocompetence (Jiang et al., 2017; Jiang et al., 2019). Recently, several genes related to gonadal development have been identified, such as *piwi* (Sun et al., 2021), *vasa* (Yan M. et al., 2013) and *foxl2* (Sun et al., 2022), and a hypothetical XX/

XY sex determination system has been proposed (Wei et al., 2021). However, the mechanisms underlying sex determination and sex differentiation in the sea cucumber remain a mystery.

Identification of sex-related genes is an effective method to reveal the genetic basis of sex determination/differentiation in the sea cucumber. The *Dmrt* gene family encodes a large family of crucial transcription factors containing one or several conserved DM domains; this family was first identified in fruit fly (*Drosophila melanogaster*) (*Arthropoda*) (Burtis and Baker, 1989). The *Dmrt* gene family is involved in various biological processes, including sex determination, sex differentiation, testicular development, and embryo development (Dong et al., 2020). To date, the *Dmrt* gene family has been identified in diverse organisms and members of the *Dmrt* gene family exhibit substantial variation in different organisms. For instance, eight *Dmrt* genes (*Dmrt1*, *Dmrt2*, *Dmrt3*, *Dmrt4*, *Dmrt5*, *Dmrt6*, *Dmrt7* and *Dmrt8*) have been identified in humans (*Homo sapiens*) (Ottolenghi et al., 2002) and mice (*Mus musculus*) (Kim et al., 2003). Six *Dmrt* genes, including *Dmrt1*, *Dmrt2*, *Dmrt3*, *Dmrt4*, *Dmrt5* and *Dmrt6,* have been identified in tuatara (*Sphenodon punctatus*), and *Dmrt1* is regarded as a crucial sex determination/differentiation gene in *S. punctatus* (Wang et al., 2006). Five *Dmrt* genes (*Dmrt1*, *Dmrt2*, *Dmrt3*, *Dmrt5* and *Dmrt6*) have been identified in fish (Dong et al., 2020). In amphibians, five *Dmrt* genes (*Dmrt1*, *Dmrt2*, *Dmrt3*, *Dmrt4* and *Dmrt5*) have been identified (Bewick et al., 2011; Watanabe et al., 2017). Due to teleost-specific whole-genome duplication (TGD) and rapid gene loss following TGD during evolution, the number of *Dmrt* genes is species-specific in teleost; there are five *Dmrt* genes in puffer fish (*Takifugu rubripes*) and zebrafish (*Danio rerio*) (Yamaguchi et al., 2006) but six *Dmrt* genes in Asian sea bass (*Lates calcarifer*), spotted gar (*Lepisosteus oculatus*) and channel catfish (*Ictalurus punctatus*) (Dong et al., 2020). Multiple *Dmrt* genes have also been identified from diverse invertebrate organisms. Seven *Dmrt* genes were identified both in Chinese mitten crab (*Eriocheir sinensis*) (*Arthropoda*), freshwater prawn (*Macrobrachium rosenbergii*) (*Arthropoda*) and mud crab (*Scylla paramamosain*) (*Arthropoda*) by screening transcriptome data using the bioinformatics method (Abayed et al., 2019; Du et al., 2019; Wan et al., 2021).

In addition to the identification of *Dmrt* gene families, several studies have focused on cloning, expression analysis and functional study of single *Dmrt* gene, especially the *Dmrt1* gene. *Dmrt1* is regarded as a conserved male-specific gene and plays a critical role in sex determination and sex differentiation in various species (Kopp, 2012). *Sox9* (SRY-related HMG box gene 9) is a member of the *SoxE* subfamily and is another essential regulator of testis determination in many organisms (Hui et al., 2021). Both *Dmrt1* and *Sox9* lay downstream of the male sex differentiation pathway, and the expression of *Dmrt1* was found to be upregulated before *Sox9* during testis development (Elzaiat et al., 2014). *Dmrt1* positively regulates the transcription of the *Sox9b* gene by directly binding to a specific cis-regulatory element (CRE) (Wei et al., 2019; Vining et al., 2021). Knockdown of the male sex-determining gene *Dmrt1* leads to decreased *Sox9* expression, while overexpression of *Dmrt1* results in upregulation of *Sox9* expression in red-eared slider turtle (*Trachemys scripta*) (Ge et al., 2017). In vertebrates, at least three members, including *Sox8*, *Sox9* and *Sox10,* have been identified in the *SoxE* subfamily, but only one ancestral *SoxE* gene has been found in all studied invertebrate species (Heenan et al., 2016). The role of *SoxE* in the specification of the neural crest and the regulation of chondrogenesis has been revealed in invertebrates (McCauley, 2008); however, little is known about its function in reproduction and the interaction between *Dmrt1* and *SoxE*.

In the current study, five members of the *Dmrt* gene family and a single *SoxE* gene were identified, and their molecular characteristics and phylogenetics were systemically analysed. Moreover, their dynamic expression patterns in different adult tissues and during different gonadal developmental stages were analysed using real-time quantitative PCR (RT-qPCR). Finally, the function and underlying interaction between *Dmrt1* and *SoxE* in the process of gonadal development were explored using RNA interference (RNAi). These findings provide valuable information for understanding the sex determination and differentiation mechanisms of Echinodermata.

# 2 Materials and methods

## 2.1 Identification of *Dmrt* gene family members and the *SoxE* gene

The published genome (taxid: 307972) data of the sea cucumber at the scaffold level (Zhang et al., 2017) was downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov/). A BLASTP search was then performed using the common conserved domain protein sequence of the DM domain (XM_030995783.1) and the High Mobility Group box (HMG box) (XP_786809.2) to screen the homologous genes of the *Dmrt* family and *SoxE*, respectively. The E-value was set to ≤ e− 5. The suspected members of the *Dmrt* gene family and *SoxE* gene were validated by blasting against the NCBI database, in order to assess the reliability of the analysis.

## 2.2 Sequence and phylogenetic analysis

The coding sequence of the *Dmrt* gene was predicted by ORF Finder online software (https://www.ncbi.nlm.nih.gov/orffinder). BioEdit software was used to perform multiple sequence alignments and visual analysis of the *Dmrt* proteins in the sea cucumber. The sequences used in this study were downloaded from the NCBI database (Supplementary Table S1). Phylogenetic analysis was performed using MEGA-7 software with bootstrapping (1,000 replicates) and the neighbour-joining (NJ) method.

## 2.3 Real-time quantitative PCR (RT-qPCR)

Adult sea cucumbers were collected from the coastal areas of Dalian, China. No endangered or protected species were involved in this study. The expression patterns of the *Dmrt* genes and *SoxE* gene in different tissues were analysed by RT-qPCR. Briefly, total RNA from different tissues, including the tube feet, testes, intestines, ovaries, stomach, longitudinal muscle, coelomocytes and respiratory tree, were extracted using the Silica Membrane-based Vacuum Pump (SV) Total RNA Isolation System (Promega Z3100). The middle part of the intestine was used for gene expression analysis. According to previously reports (Sun et al., 2021), the gonads of *A. japonicus* are classified into four stages, including the early growing stage (Stage 1), growing stage (Stage 2), mature stage (Stage 3) and post-spawning stage (Stage 4). After gonadal histology analysis, gonadal tissues at

different development stages were collected to examine the dynamic expression changes in *Dmrt1* and *SoxE*. Gonadal tissues at Stage 2 were used to detect tissue expression specificity.

RT-qPCR experiments were performed in 20-μl reactions containing 10 μL of Fast Start Essential DNA Green Master (Roche, Mannheim, Germany), 6.4 μL of sterile water, 0.8 μL of each 10 mM primer, and 2 μL of cDNA. The protocol was as follows: 95°C (10 min) for heat denaturing; then, 40 cycles of 95°C (15 s) and 60°C (1 min). The efficiency of primers discovery through standard curve formulation reached 98%. The housekeeping gene *NADH* was selected as the internal reference gene (Sun et al., 2021). Data were performed from three independent experiments. Each sample was analyzed in triplicates, and the relative expression levels of target genes were calculated with the $2^{-\Delta\Delta Ct}$ method. For statistical analysis, one-way ANOVA was calculated with SPSS software after the normal distribution and homogeneity of variance test (SPSS Inc.), and a probability ($p$) of ≤0.05 was considered statistically significant. The primers used in this study were designed using online software (http://biotools.nubic.northwestern.edu/OligoCalc.html) (Supplementary Table S2).

## 2.4 RNA interference

Gene-specific dsRNAs for the *Dmrt1* and *SoxE* genes were designed using an online platform (https://www.dkfz.de/signaling/e-rnai3/). DsRNAs were synthesised using a T7 RiboMAX™ Express RNAi System (Promega), according to the manufacturer's protocol. Three different dsRNAs were designed for each gene to ensure interference efficiency. The genetic sex of each sea cucumber was determined according to a previous report (Wei et al., 2021). In briefly, the tube feet tissues were used to extract genomic DNA by alkaline lysis method, then male-specific primers were used to amplify the male specific DNA marker, one specific band was amplified in males, but not in females. For the *Dmrt1* RNAi experiment, 24 male sea cucumbers with an average weight of 100 ± 20 g were selected and were randomly divided into two groups: *Dmrt1*-knockdown ($n$ = 12) and control ($n$ = 12). For the *SoxE* RNAi experiment, 24 male sea cucumbers with an average weight of 60 ± 10 g were selected. The 24 males were divided into two groups: *SoxE*-knockdown ($n$ = 12) and control group ($n$ = 12). The injection experiments were performed as previously reported with slight modification (Sun et al., 2021). Specifically, before injecting, three gene-specific dsRNAs were mixed, then 100 μg of gene-specific dsRNA per 50 g body weight was injected into each sea cucumber in the knockdown group. Meanwhile, sea cucumbers in the control group were injected with GFP dsRNA at the same concentration. For the *Dmrt1* RNAi experiment, three sea cucumbers from each group were removed randomly on the third day, seventh day and 10th day following the first injection. The gonads were surgically sampled for future research. In addition, repeat injections were performed on the third, sixth and ninth days following the first injection. For the *SoxE* RNAi experiment, three sea cucumbers from each group were removed randomly on the third, 7th and 23rd day following the first injection. The gonads were surgically sampled for future research. In addition, repeat injections were performed every 3 days after the first injection.

## 2.5 Gonadal histology

Gonadal tissues were surgically removed and fixed in 4% paraformaldehyde (PFA) at 4°C overnight. Then, they were washed three times in phosphate-buffered solution (PBS) and balanced in 30% sucrose at room temperature for 2 h. Next, the treated tissues were embedded in optimal cutting temperature (O.C.T) compound and frozen sections (5 μm) were cut and stained with haematoxylin/eosin. Digital photos were acquired under a Leica DM4B microscope

# 3 Results

## 3.1 Identification and characterization of *Dmrt* and *SoxE* genes in *A. japonicus*

Genomic information for *A. japonicus* is available from NCBI (assembly ASM275485v1) (Zhang et al., 2017). To identify the *Dmrt* and *SoxE* genes in *A. japonicus*, the protein sequences of the conserved DM domain and HMG box of *S. purpuratus* (XM_030995783.1 and XP_786809.2) were used as the queries to perform a BLASTP search, respectively. Five predicted *Dmrt* genes (PIK43860.1, PIK34621.1, PIK44057.1, PIK33706.1 and PIK41536.1) were identified. Each *Dmrt* gene contained a complete coding DNA sequence (CDS) with lengths of 1,029, 651, 945, 507 and 849 bp, respectively. All identified *Dmrt* proteins had a DM domain (Supplementary Figure S1A), but had a different number of exons. PIK41536.1, PIK43860.1 and PIK34621.1 contained just a single intron, PIK44057.1 had two introns and PIK33706.1 had no introns (Supplementary Figure S1B). The identity between *Dmrt* cDNA sequences is about 32 %. Likewise, a predicted *SoxE* gene encoding a protein of 459 amino acids was identified *AjSoxE*. As expected, there was a 71-amino acid HMG box domain in the protein sequence (Supplementary Figure S1C).

To define these five *Dmrt* genes, 40 *Dmrt* genes from 17 species were downloaded from the NCBI database and multiple protein sequence alignments and phylogenetic analyses were performed. The protein sequences of the *Dmrt* genes exhibiting low similarity between species (4.94%–57.01%). However, the protein sequence of DM domain was highly conserved such that the similarity with other species was generally more than 60% (Supplementary Figures S2–S5). The phylogenetic analyses indicated that the *Dmrt* family genes can be classified into five branches (*Dmrt*1, -2, -3, -4 and -5). PIK41536.1 and PIK43860.1 were clustered into the *Dmrt1* and *Dmrt2* branches, and were named *AjDmrt1* and *AjDmrt2*, respectively. PIK34621.1 and PIK44057.1 were both clustered with *Dmrt3* homologs from other species, so were named *AjDmrt3a* and *AjDmrt3b*, respectively. The deduced amino acid sequences of PIK33706.1 were clustered into the *Dmrt5* branch and were defined as *AjDmrt5* (Figure 1A). Surprisingly, none of the *AjDmrt* genes clustered with *Dmrt4* homologs from other species' branches. The topology of clades of *Dmrt1*, *Dmrt2*, *Dmrt3* and *Dmrt5* were basically consistent with the known taxonomic relationships among these species. Moreover, the phylogenetic analyses showed that *AjSoxE* was firstly grouped with *SoxE* in invertebrate species, then clustered into *Sox8*, *Sox9* and *Sox10* branches in jawed vertebrates (Figure 1B). This phylogenetic tree further confirms that the ancestral *SoxE* gene is replicated in vertebrates and has produced at least three members (*Sox8*, *Sox9*, and *Sox10*).

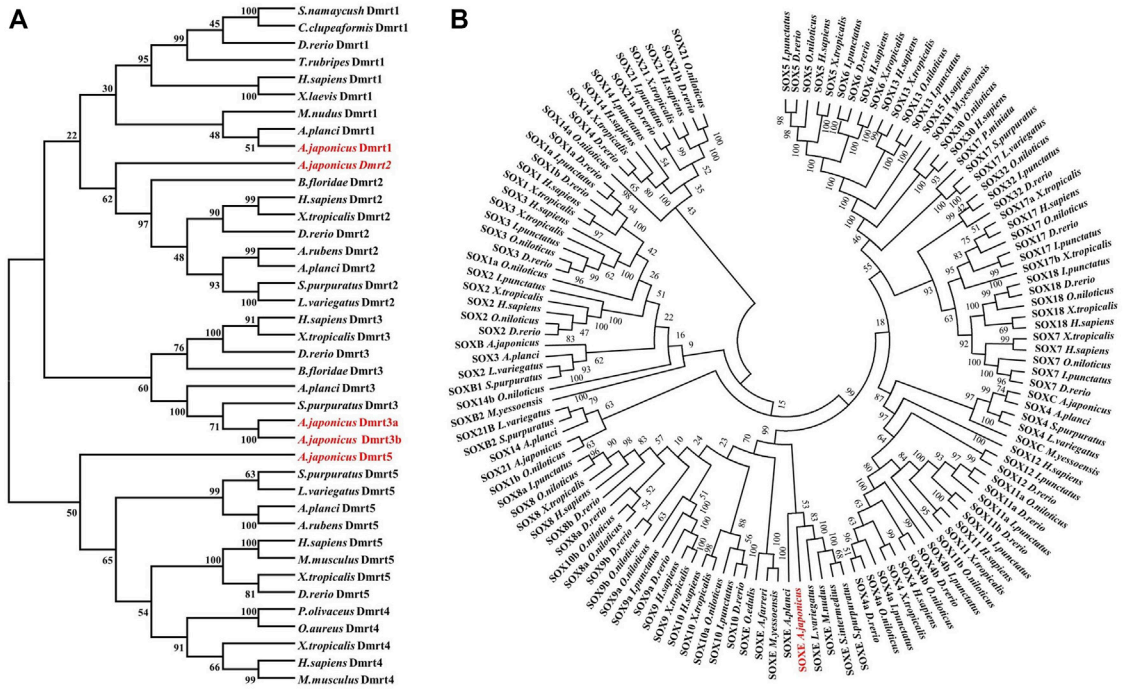



FIGURE 1
Phylogenetic analysis of the Dmrt and SOX protein family in *Apostichopus japonicus* **(A)** Dmrt.**(B)** SoxE. The phylogenetic tree was constructed with MEGA version 7.0 (1,000 replicates) by maximum likelihood analysis.

## 3.2 Expression of *Dmrts* and *soxE* in adult tissues in *A. japonicus*

In order to reveal the expression differences, the transcripts of all five *AjDmrt* genes and the *SoxE* gene in adult tissues were examined by RT-qPCR. Specific primers of five *AjDmrt* genes were designed in the parts with the greatest difference in the nucleic acid sequence. *AjDmrt1* was expressed predominantly in the testes, only low *AjDmrt1* expression was detected in other tissues, including the ovaries, stomach, tube feet, longitudinal muscle, respiratory tree, and coelomocytes (Figure 2A). In contrast, *AjDmrt2* was higher expressed in the tube feet (Figure 2B). *AjDmrt3a* and *AjDmrt3b* were expressed predominantly in the coelomocytes and tube feet, with low expression of *AjDmrt3a* and *AjDmrt3b* detected in other tissues, including the intestines, stomach, longitudinal muscle, ovaries, and respiratory tree (Figures 2C,D). The expression pattern of *AjDmrt5* in adult tissue was similar to that of *AjDmrt2*, with predominant expression in the tube feet and slight expression in the other analysed tissues (Figure 2E). Strikingly, the *SoxE* gene had the highest expression in the tube feet. Its expression differed significantly between the male and female gonads, such that a large number of transcripts were detected in the ovaries while *SoxE* was nearly undetectable in the testes (Figure 2F).

## 3.3 Expression profile of *Dmrt1* and *SoxE* during gonadal development in *A. japonicus*

Considering that the expression of *Dmrt1* and *SoxE* differed significantly between the ovaries and testes in *A. japonicus*, gonadal tissues at different development stages were collected to examine the dynamic expression changes in *Dmrt1* and *SoxE*. After gonadal histology analysis, gonadal tissues of four different development stages (i.e., the early growing stage (Stage 1), growing stage (Stage 2), mature stage (Stage 3) and post-spawning stage (Stage 4)) were obtained from male and female individuals, respectively. The expression of *Dmrt1* was continuously maintained at a high level during the spermatogenesis process and was significantly downregulated during Stage 2 in the testes (Figure 3A). *SoxE* expression in the testes at the corresponding stage was much lower than that of *Dmrt1*, and there were no significant differences between Stage 1, Stage 2 and Stage 3. Along with the occurrence of spermatogenesis, the expression level of *SoxE* increased up to 6.3-fold during Stage 4 against to Stage1 (Figure 3B). Meanwhile, a large amount of *SoxE* transcript was detected in Stage 1 ovary tissue, but the relative expression level continuously decreased together with the oogenesis development process, reaching the lowest in Stage 4 (Figure 3C).

## 3.4 Knockdown of *Dmrt1* by RNAi in male *A. japonicus*

RNAi was further utilised to examine the underlying function of *Dmrt1* in testes development in male *A. japonicus*. As shown in Figure 4A, the expression levels of *Dmrt1* significantly decreased, down 40.85% and 47.80% in the knockdown group compared to the control group at 3 days post-injection (dpi) and 7 dpi, respectively. On the 10 dpi, *Dmrt1* expression was only 3.31% of that in the control group. Compared to the control group, the transcripts of *SoxE* were

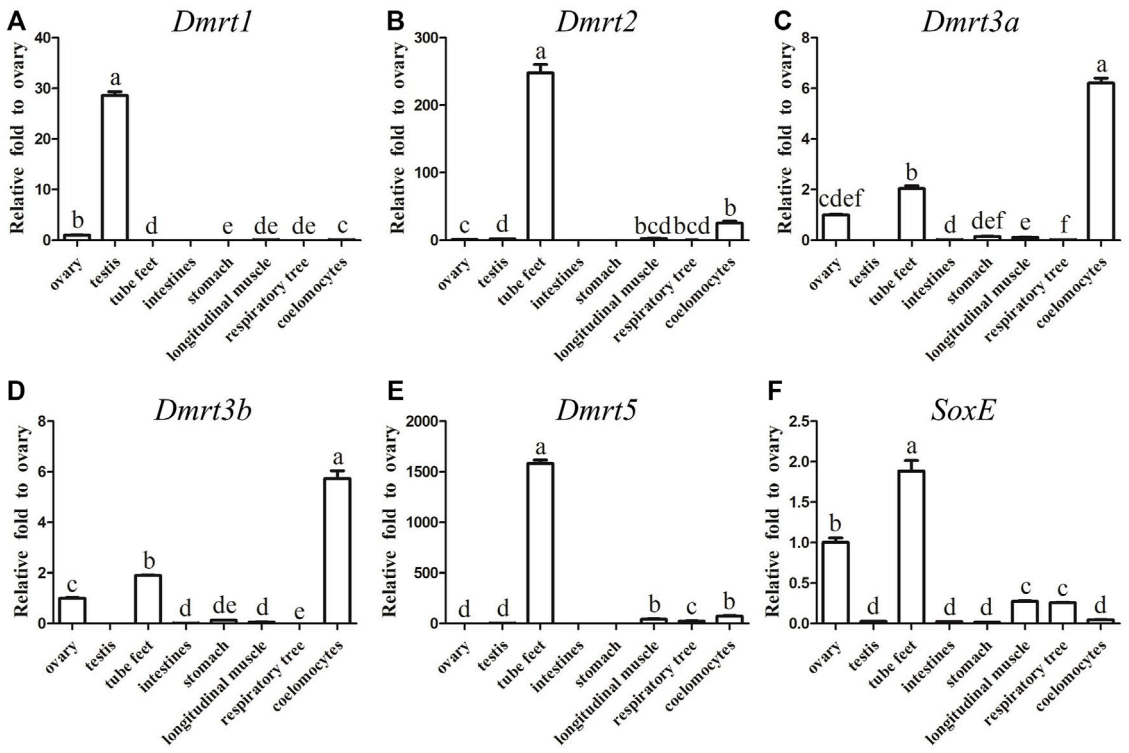



FIGURE 2
RT-qPCR analysis of *Dmrts* and *SoxE* Expression in adult tissues. NADH was used as internal control **(A)** Dmrt1. **(B)** Dmrt2. **(C)** Dmrt3a. **(D)** Dmrt3b. **(E)** Dmrt5. **(F)** SoxE. Data were performed from three independent experiments. Each bar represents mean ± SD. Different letters indicate significant differences between mean values ($p \leq 0.05$), whilst shared letters indicate no significant difference.

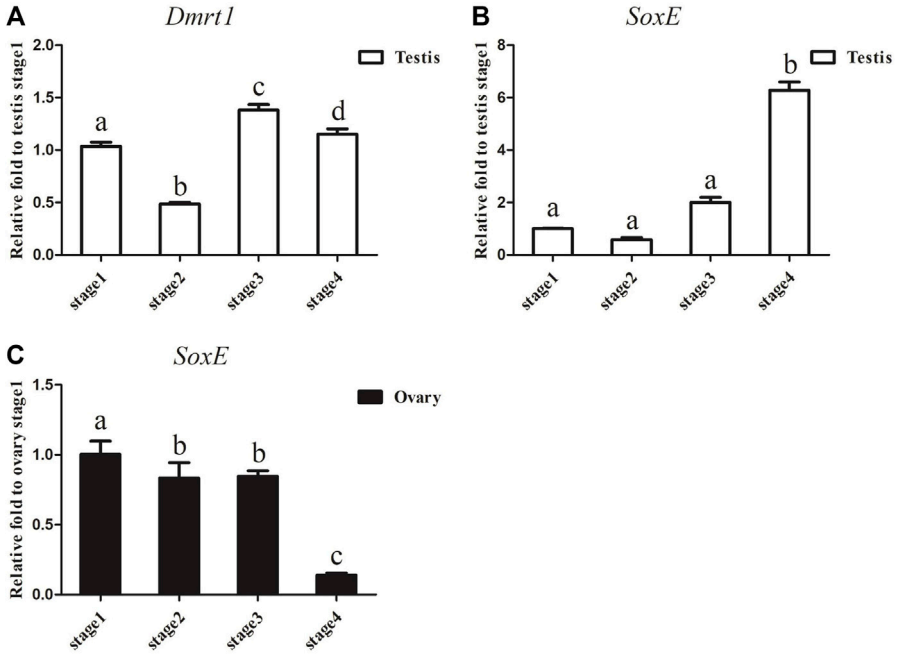



FIGURE 3
*Dmrt1* and *SoxE* expression in gonad. **(A)** RT-qPCR analysis of *Dmrt1* expression in testis at different development stages. **(B)** RT-qPCR analysis of *SoxE* expression in testis at different development stages. **(C)** RT-qPCR analysis of *SoxE* expression in ovary at different development stages. NADH was used as internal control. Data were performed from three independent experiments. Each bar represents mean ± SD. Different letters indicate significant differences between mean values ($p \leq 0.05$), whilst shared letters indicate no significant difference.

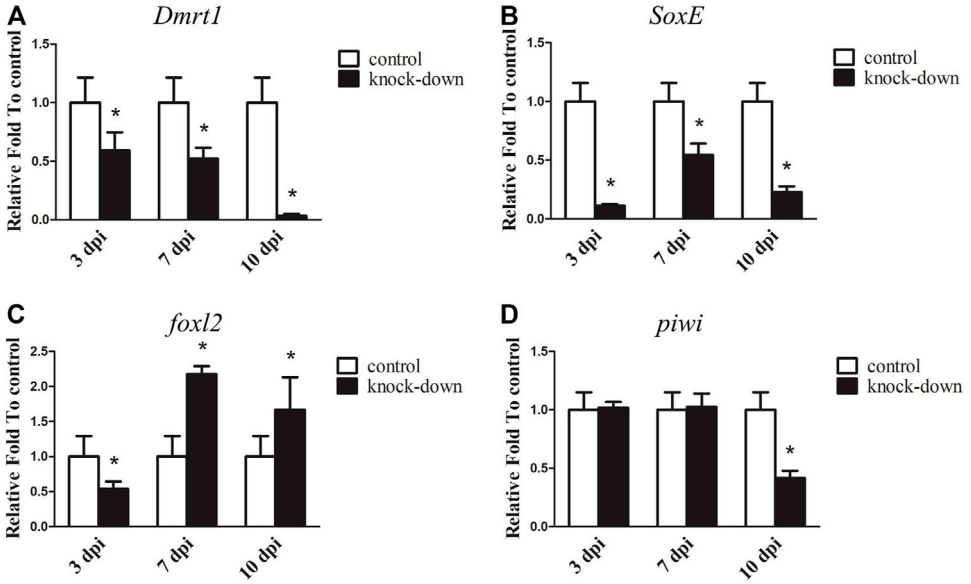



FIGURE 4
RNA interference (RNAi) of *Dmrt1* in testis. RT-qPCR detected mRNA levels of *Dmrt1* and other sex related genes after *Dmrt1* RNAi. **(A)** Dmrt1. **(B)** *SoxE*. **(C)** *foxl2*. **(D)** *piwi*. NADH was used as internal control. Data were performed from three independent experiments. Each bar represents mean ± standard deviation (SD). Asterisks (*) indicate significant differences ($p \leq 0.05$) between knock-down and control. dpi: days post-injection.

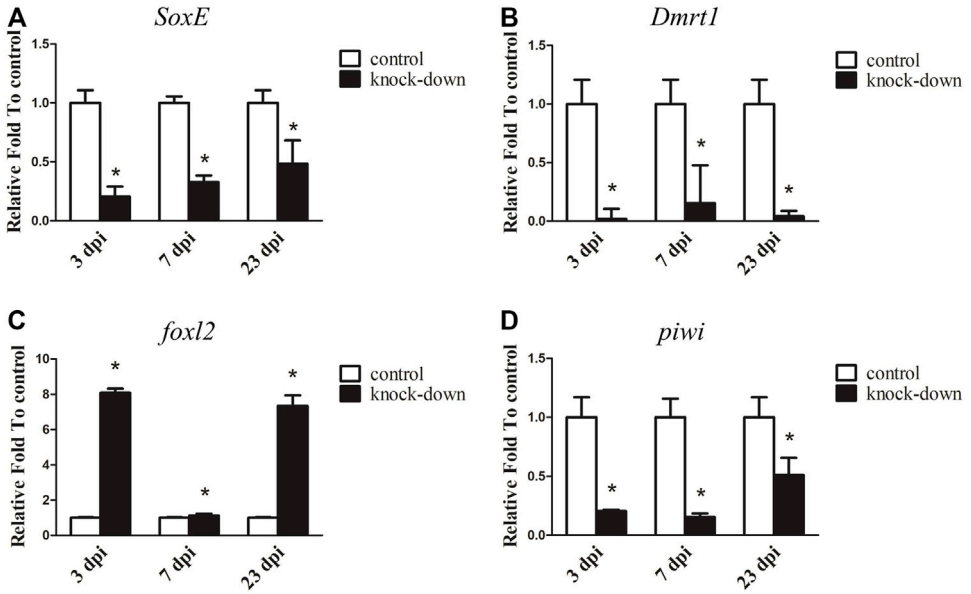



FIGURE 5
RNA interference (RNAi) of *SoxE* in testis. RT-qPCR detected mRNA levels of *SoxE* and other sex related genes after *SoxE* RNAi. **(A)** *SoxE*. **(B)** *Dmrt1*. **(C)** *foxl2*. **(D)** *piwi*. NADH was used as internal control. Data were performed from three independent experiments. Each bar represents mean ± standard deviation (SD). Asterisks (*) indicate significant differences ($p \leq 0.05$) between knock-down and control. dpi: days post-injection.

reduced in the knockdown group by 88.65% at 3 dpi, 45.52% at 7 dpi and 77.10% at 10 dpi (Figure 4B). Moreover, the dynamic expression changes in *foxl2* and *piwi*, which have been reported to be involved in ovary differentiation and germ cell development, respectively (Ottolenghi et al., 2005; Gonzalez et al., 2021), were investigated. Interestingly, the expression level of *foxl2* was significantly downregulated (about 46.17%) in the knockdown group at 3 dpi, but sharply increased 1.67–2.17-fold at 7 dpi and 10 dpi (Figure 4C). The expression level of *piwi* did not significantly differ between the knockdown and control groups at 3 dpi and 7 dpi, but was

significantly reduced by 58.36% at 10 dpi in knockdown adults (Figure 4D). Although the expression levels of genes involved in testis differentiation, ovary differentiation and germ cell development showed dynamic changes after knockdown of *Dmrt1* by RNAi, no obvious histological changes and no apoptosis cells were observed in the knockdown testes (Supplementary Figure S6).

### 3.5 knockdown of *SoxE* by RNAi in male *A. japonicus*

Since *SoxE* has been reported to play a conserved role in sexual differentiation, and it shows a sexually dimorphic expression pattern in the gonads of *A. japonicus,* its function during testis development was investigated by RNAi in an adult male individual. As shown in Figure 5A, the transcript level of *SoxE* was significantly decreased by 79.46% with *SoxE* knockdown compared to the control at 3 dpi. Its expression was continuously downregulated at 7 dpi and 23 dpi. In addition, the expression of *Dmrt1* was sharply decreased to almost an undetectable level in the knockdown testes (Figure 5B). Conversely, the expression level of *foxl2* was increased about 8.09-fold and 7.34-fold in the knockdown sample compared to the control sample at 3 dpi and 23 dpi, respectively. However, the expression level of *foxl2* was not significantly different compared to the control at 7 dpi (Figure 5C). The transcripts of *piwi* were significantly decreased by 79.42%, 84.60% and 48.87% at 3 dpi, 7 dpi and 23 dpi, respectively (Figure 5D ). However, no obvious histological changes or apoptosis cells were observed in the knockdown testes (Supplementary Figure S7).

## 4 Discussion

With the development of sequencing technology, genome-wide investigations of gene families have become common, and several gene families, including the *Dmrt* gene family (Zafar et al., 2019), *Sox* gene family (Yang et al., 2020), *DEAD-box* helicase family (Xu et al., 2021), *NRAMP* gene family (Tian et al., 2021), *TRAF* gene family (Zhang et al., 2020), *TGF-beta* gene family (Cao et al., 2018) and *TLR* gene family (Habib et al., 2021) have been identified in various species. The origin of the *Dmrt* gene family can be traced back to *coelenterates* in the animal kingdom (Bellefroid et al., 2013; Wexler et al., 2014), and their composition, expression, and function show species-specificity. Here, *Dmrt3* genes were duplicated in the *A. japonicus* genome and *Dmrt3a* and *Dmrt3b* were produced, whereas *Dmrt4* was lost. In other species, *Dmrt4* and *Dmrt5* were clustered into a major branch (Wang et al., 2012; Dong et al., 2020), indicating that these genes possibly originated from a common ancestor of *Dmrt*. This loss or duplicated gene phenomenon possibly resulted from independent gene duplication (loss) during the evolution of the sea cucumber. In most vertebrates, the *Dmrt1–Dmrt3–Dmrt2*(2a) gene cluster is highly conserved and plays a key role in sex differentiation (Johnsen and Andersen., 2012; Dong et al., 2020). Since the existing sea cucumber genome is not assembled to the chromosome level (Zhang et al., 2017; Li et al., 2018), we failed to identify the *Dmrt1–Dmrt3–Dmrt2*(2a) conserved gene cluster (Supplementary Figure S8). Hence, a chromosome-level assembly of the sea cucumber (*A. japonicus*) should be sequenced to further elucidate the linkages among the conserved gene cluster and better understand the sex-determination mechanism.

Two different expression patterns of *Dmrt1* have been identified, including testes-specific expression and sexually dimorphic expression. In most cases, *Dmrt1* is specifically expressed in the testes and plays essential roles in male-sex determination and testicular differentiation. Knockdown or knockout of *Dmrt1* leads to male-to-female sex reversal (Paul-Prasanth et al., 2006; Luo et al., 2015; Chakraborty et al., 2016). However, there are some situations in which *Dmrt1* shows a sexually dimorphic expression pattern, with slight expression in the ovaries and abundant expression in the testes; for example, in red-eared slider turtle (*Trachemys scripta*) (Ge et al., 2017). In sea cucumber, *Dmrt1* shares a similar expression pattern with the red-eared slider turtle, with higher expression in the testes than in the ovaries. In addition, its expression was found to be highly expressed throughout spermatogenesis, but declined at the growing stage (Figure 3A). *Dmrt1* expression was decreased at spermiation stage in comparison with growing and maturing stages in *Oncorhynchus mykiss* and *Odontesthes bonariensis* (Marchand et al., 2000; Fernandino et al., 2006). These differences may be due to the different reproductive strategies and different testis structure between sea cucumber and fish, and should be investigated in the future. Curiously, there appears to be greater expression of *SoxE* in the ovaries than the testes at the same developmental stages. Both *Dmrt1* and *Sox9* are necessary for male sexual development in several species, knockdown (off) of *SoxE* (*Sox9*) and *Dmrt1* leads to the disappearance of the male marker, ectopic expression of the ovarian regulator and the formation of an ovary-like structure (Barrionuevo et al., 2006; Kim et al., 2007; Krentz et al., 2009; Cui et al., 2017). In Nile tilapia knockdown of *Dmrt1* led to the downregulation of *Sox9b,* and two potential cis-regulatory elements (CREs) for the *Dmrt1* transcription factor were identified in the promoter of *Sox9b* (Wei et al., 2019). Loss of *DMRT1* will inhibit *SOX9* expression and activates *FOXL2* transcription factor, even can reprogram granulosa cells into Sertoli cells in mice (Matson et al., 2011). In most species, sex is determined by male regulatory gene network in which *Sry* activates *Sox9* and a female network involving *WNT/b-catenin* signalling (Matson et al., 2011), and *Dmrt1* recruits *Sox9* to reprogram sexual cell fate(Lindeman et al., 2021). Although no obvious histological changes were observed in *SoxE* and *Dmrt1* knockdown testes in the current study, the expression level of the ovarian regulator *foxl2* was significantly upregulated at 3dpi and 23dpi, indicating that *SoxE* and *Dmrt1* may be closely related to the development and maintenance of the testes. Unexpectedly, *foxl2* was significantly downregulated at 7dpi after *SoxE*-knockdown. Although RNAi provides a useful approach to study gene function in non-model species (Yang et al., 2017; Yu et al., 2014; Zhang et al., 2019), it is indeed a big challenge to ensure considerable knockdown efficiency among individuals especially in echinoderms with open-tube circulation. Hence, it is of great value to evaluate the functions of *SoxE* and *Dmrt1* in testes differentiation through gene knockout biotechnologies.

Apart from *Dmrt1*, which is highly expressed in the testes, *Dmrt2*, *3a*, *3b* and *5*, as well as *SoxE,* showed significantly higher expression in tube feet than in other adult tissues. Current studies have shown that the tube feet of the sea cucumber has a wide range of functions, including locomotion, feeding, chemoreception and light sensitivity respiration (Wang et al., 2014; Sun et al., 2013). According to previous reports, the expression pattern of *Dmrt2* (*2a*, *2b*) are species-specific, with expression primarily in both the ovaries and testes (Yamaguchi et al., 2006; Sheng et al., 2014; Su

et al., 2015). In vertebrates, *Dmrt3* shows high expression in the testes and nervous system; accordingly, it has been speculated to be involved in the development of nerves and germ cells (Yamaguchi et al., 2006; Li et al., 2008; Dong et al., 2010). In contrast, large transcripts of *Dmrt3a* and *Dmrt3b* have been detected in coelomocytes, indicating that they might contribute to the regulation of immune functions. Surprisingly, the expression of *SoxE* was higher in the ovaries than the testes. In vertebrates, the transcription factor *Sox9* is highly expressed in developing male gonadal ridges and serves as the master regulator of Sertoli cell differentiation, playing a crucial role in sex determination (Kent et al., 1996; Jakob and Lovell-Badge., 2011). In invertebrates, *SoxE* has been detected in both the ovaries and testes and is mainly located in oocytes, spermatocytes and Sertoli cells (Santerre et al., 2014; Li et al., 2016; Yu et al., 2017). This indicates that *SoxE* may be involved in ovary development and differentiation as well as in the development of the testis in invertebrates.

## 5 Conclusion

In conclusion, this study identified and characterized the *SoxE* gene and *Dmrt* gene family in *A.japonicus* for the first time. A total of five *Dmrt* genes (*Dmrt1*, *2*, *3a*, *3b* and *5*) were identified and classified into four families based on phylogenetic analyses. Expression profile analyses demonstrated that the express patterns of *Dmrt* genes and the *SoxE* gene. Among them, *Dmrt1* was higher expressed in the testes with low expression in the ovaries. Knocking-down of *Dmrt1* led to the downregulation of *SoxE* expression and upregulation of the ovarian regulator *foxl2* in the testes. This indicates that *Dmrt1* may be a positive regulator of *SoxE* and may play a role in the development of the testes in the sea cucumber. In contrast, *SoxE* was more highly expressed in the ovaries than the testes, and the transcript level of *Dmrt1* was significantly decreased while *foxl2* expression was sharply increased in *SoxE* knockdown testes. This study provides a comprehensive overview of the *Dmrt* gene family and contributes to a better understanding of the role of the *Dmrt1* and *SoxE* genes in the sex determination and differentiation mechanisms of the sea cucumber.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

## Author contributions

"Conceptualization, Z-HS and Y-QC; methodology, B-ZL; investigation, B-ZL, J-JC and W-YS; resources, Z-HS and Y-QC; data curation Z-LH and B-ZL; writing—original draft preparation, Z-HS and B-ZL; writing—review and editing, Z-HS and B-ZL; funding acquisition, Z-HS and Y-QC; All authors have read and agreed to the published version of the manuscript."

## Funding

This work was supported by the National Key R&D Program of China (grant number 2018YFD0900201) and Liaoning Province "Xingliao Talents Plan" project (Grant no. XLYC2002107).

## Conflict of interest

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

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1097825/full#supplementary-material

**SUPPLEMENTARY FIGURE S1**
Gene structure of *Dmrt* and *SoxE*. **(A)** Conserved DM domain diagram of *Dmrts*. **(B)** Schematic diagram of the gene structure of *Dmrts*. **(C)** Conserved HMG domain diagram of *SoxE*.

**SUPPLEMENTARY FIGURE S2**
Amino acid alignments of Dmrt1 with the homologs from other species. **(A)** Full length of amino acid alignments of Dmrt1. **(B)** DM domain alignments of Dmrt1. The identities relative to are exhibited at the end of each sequence. The boule frame indicates the conserved position of the DM domain.

**SUPPLEMENTARY FIGURE S3**
Amino acid alignments of Dmrt2 with the homologs from other species. **(A)** Full length of amino acid alignments of Dmrt2. **(B)** DM domain alignments of Dmrt2. The identities relative to are exhibited at the end of each sequence. The boule frame indicates the conserved position of the DM domain.

**SUPPLEMENTARY FIGURE S4**
Amino acid alignments of Dmrt3 with the homologs from other species. **(A)** Full length of amino acid alignments of Dmrt3. **(B)** DM domain alignments of Dmrt3. The identities relative to are exhibited at the end of each sequence. The boule frame indicates the conserved position of the DM domain.

**SUPPLEMENTARY FIGURE S5**
Amino acid alignments of Dmrt5 with the homologs from other species. **(A)** Full length of amino acid alignments of Dmrt5. **(B)** DM domain alignments of Dmrt5. The identities relative to are exhibited at the end of each sequence. The boule frame indicates the conserved position of the DM domain.

**SUPPLEMENTARY FIGURE S6**
Histological examination of testes after RNAi of *Dmrt1*. Bars=50 μm.

**SUPPLEMENTARY FIGURE S7**
Histological examination of testes after RNAi of *soxE*. Bars=50 μm.

**SUPPLEMENTARY FIGURE S8**
Gene synteny of Dmrt3 with the homologues from other species. Conserved gene blocks are represented in matching colors, and transcription orientations are indicated by arrows.

# References


Abayed, F. A. A., Manor, R., Aflalo, E. D., and Sagi, A. (2019). Screening for *dmrt* genes from embryo to mature *Macrobrachium rosenbergii* prawns. *General Comp. Endocrinol.* 282, 113205. doi:10.1016/j.ygcen.2019.06.009

Barrionuevo, F., Bagheri-Fam, S., Klattig, J., Kist, R., Taketo, M. M., Englert, C., et al. (2006). Homozygous inactivation of *Sox9* causes complete XY sex reversal in mice. *Biol. Reprod.* 74 (1), 195–201. doi:10.1095/biolreprod.105.045930

Bellefroid, E. J., Leclère, L., Saulnier, A., Keruzore, M., Sirakov, M., Vervoort, M., et al. (2013). Expanding roles for the evolutionarily conserved *dmrt* sex transcriptional regulators during embryogenesis. *Cell. Mol. life Sci.* 70 (20), 3829–3845. doi:10.1007/s00018-013-1288-2

Bewick, A. J., Anderson, D. W., and Evans, B. J. (2011). Evolution of the closely related, sex-related genes *DM-W* and *dmrt1* in African clawed frogs (*Xenopus*). *Evol. Int. J. Org. Evol.* 65 (3), 698–712. doi:10.1111/j.1558-5646.2010.01163.x

Burtis, K. C., and Baker, B. S. (1989). Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* 56 (6), 997–1010. doi:10.1016/0092-8674(89)90633-8

Cao, L., Xie, B. J., Cui, B. B., Chen, B. B., Ma, Z. S., Chen, X. B., et al. (2018). Genome-wide feature analysis of the sequence-specific recognition in intermolecular interaction between *tgf-β* pathway dna and *meg3 lncrna* in human cancer. *J. Chemom.* 32, e3042. doi:10.1002/cem.3042

Chakraborty, T., Zhou, L. Y., Chaudhari, A., Iguchi, T., and Nagahama, Y. (2016). *Dmy* initiates masculinity by altering *Gsdf/Sox9a2/Rspo1* expression in medaka (*Oryzias latipes*). *Sci. Rep.* 6 (1), 19480–19511. doi:10.1038/srep19480

Chen, J., and Chang, Y. (2015). Sea cucumber aquaculture in China. *Echinoderm Aquac.* 14, 317–330. doi:10.1002/9781119005810.ch14

Cui, Z., Liu, Y., Wang, W., Wang, Q., Zhang, N., Lin, F., et al. (2017). Genome editing reveals *dmrt1* as an essential male sex-determining gene in Chinese tongue sole (*Cynoglossus semilaevis*). *Sci. Rep.* 7 (1), 42213–42310. doi:10.1038/srep42213

Dong, X. L., Chen, S. L., and Ji, X. S. (2010). Molecular cloning and expression analysis of *Dmrt3* gene in half-smooth tongue sole (*Cynoglossus semilaevis*). *J. Fish. China.* 34, 829–835. doi:10.1007/s10695-018-0472-6

Dong, J. J., Li, J., Hu, J., Sun, C. F., Tian, Y. Y., Li, W. H., et al. (2020). Comparative genomics studies on the *dmrt* gene family in fish. *Front. Genet.* 11, 563947. doi:10.3389/fgene.2020.563947

Du, J., Liu, Y., Song, C. W., and Cui, Z. X. (2019). Discovery of sex-related genes from embryonic development stage based on transcriptome analysis in *Eriocheir sinensis*. *Gene* 710, 1–8. doi:10.1016/j.gene.2019.05.021

Elzaiat, M., Jouneau, L., Thépot, D., Klopp, C., Allais-Bonnet, A., Cabau, C., et al. (2014). High-throughput sequencing analyses of XX genital ridges lacking *FOXL2* reveal *DMRT1* up-regulation before *SOX9* expression during the sex-reversal process in goats. *Biol. Reprod.* 91 (6), 153–161. doi:10.1095/biolreprod.114.122796

Fernandino, J. I., Guilgur, L. G., and Somoza, G. M. (2006). *Dmrt1* expression analysis during spermatogenesis in pejerrey, *Odontesthes bonariensis*. *Fish Physiol. Biochem.* 32 (3), 231–240. doi:10.1007/s10695-006-9005-9

Ge, C. T., Ye, J., Zhang, H. Y., Zhang, Y., Sun, W., Sang, Y. P., et al. (2017). *Dmrt1* induces the male pathway in a turtle species with temperature-dependent sex determination. *Development* 144 (12), 2222–2233. doi:10.1242/dev.152033

Gonzalez, L. E., Tang, X., and Lin, H. (2021). Maternal Piwi regulates primordial germ cell development to ensure the fertility of female progeny in *Drosophila*. *Genetics* 219 (1), iyab091. doi:10.1093/genetics/iyab091

Habib, Y. J., Wan, H. F., Sun, Y. L., Shi, J. L., Yao, C. J., Lin, J. M., et al. (2021). Genome-wide identification of *toll-like receptors* in Pacific white shrimp (*litopenaeus vannamei*) and expression analysis in response to *vibrio parahaemolyticus* invasion. *Aquaculture* 532, 735996. doi:10.1016/j.aquaculture.2020.735996

Heenan, P., Zondag, L., and Wilson, M. J. (2016). Evolution of the *Sox* gene family within the chordate phylum. *Gene* 575 (2), 385–392. doi:10.1016/j.gene.2015.09.013

Hui, H. B., Xiao, L., Sun, W., Zhou, Y. J., Zhang, H. Y., and Ge, C. T. (2021). *Sox9* is indispensable for testis differentiation in the red-eared slider turtle, a reptile with temperature-dependent sex determination. *Zool. Res.* 42 (6), 721–725. doi:10.24272/j.issn.2095-8137.2021.136

Jakob, S., and Lovell-Badge, R. (2011). Sex determination and the control of *Sox9* expression in mammals. *FEBS J.* 278 (7), 1002–1009. doi:10.1111/j.1742-4658.2011.08029.x

Jiang, J. W., Zhou, Z. C., Dong, Y., Gao, S., Sun, H. J., Chen, Z., et al. (2017). Comparative analysis of immunocompetence between females and males in the sea cucumber *Apostichopus japonicus*. *Fish Shellfish Immunol.* 63, 438–443. doi:10.1016/j.fsi.2017.02.038

Jiang, J. W., Zhao, Z. L., Pan, Y. J., Dong, Y., Gao, S., Li, S. L., et al. (2019). Gender specific differences of immune competence in the sea cucumber *Apostichopus japonicus* before and after spawning. *Fish Shellfish Immunol.* 90, 73–79. doi:10.1016/j.fsi.2019.04.051

Johnsen, H., and Andersen, Ø. (2012). Sex dimorphic expression of five *dmrt* genes identified in the Atlantic cod genome. The fish-specific *dmrt2b* diverged from *dmrt2a* before the fish whole-genome duplication. *Gene* 505 (2), 221–232. doi:10.1016/j.gene.2012.06.021

Kent, J., Wheatley, S. C., Andrews, J. E., Sinclair, A. H., and Koopman, P. (1996). A male-specific role for *SOX9* in vertebrate sex determination. *Development* 122 (9), 2813–2822. doi:10.1242/dev.122.9.2813

Kim, S., Kettlewell, J. R., Anderson, R. C., Bardwell, V. J., and Zarkower, D. (2003). Sexually dimorphic expression of multiple doublesex-related genes in the embryonic mouse gonad. *Gene Expr. Patterns* 3 (1), 77–82. doi:10.1016/S1567-133X(02)00071-6

Kim, S., Bardwell, V. J., and Zarkower, D. (2007). Cell type-autonomous and non-autonomous requirements for *Dmrt1* in postnatal testis differentiation. *Dev. Biol.* 307 (2), 314–327. doi:10.1016/j.ydbio.2007.04.046

Kopp, A. (2012). *Dmrt* genes in the development and evolution of sexual dimorphism. *Trends Genet.* 28 (4), 175–184. doi:10.1016/j.tig.2012.02.002

Krentz, A. D., Murphy, M. W., Kim, S., Cook, M. S., Capel, B., Zhu, R., et al. (2009). The DM domain protein DMRT1 is a dose-sensitive regulator of fetal germ cell proliferation and pluripotency. *Proc. Natl. Acad. Sci.* 106 (52), 22323–22328. doi:10.1073/pnas.0905431106

Li, Q., Zhou, X., Guo, Y., Shang, X., Chen, H., Lu, H., et al. (2008). Nuclear localization, DNA binding and restricted expression in neural and germ cells of zebrafish Dmrt3. *Biol. Cell.* 100, 453–463. doi:10.1042/BC20070114

Li, Y., Zhang, L., Sun, Y., Ma, X., Wang, J., Li, R., et al. (2016). Transcriptome sequencing and comparative analysis of ovary and testis identifies potential key sex-related genes and pathways in scallop *Patinopecten yessoensis*. *Mar. Biotechnol.* 18 (4), 453–465. doi:10.1007/s10126-016-9706-8

Li, Y., Wang, R., Xun, X., Wang, J., Bao, L., Thimmappa, R., et al. (2018). Sea cucumber genome provides insights into saponin biosynthesis and aestivation regulation. *Cell Discov.* 4 (1), 29–17. doi:10.1038/s41421-018-0030-5

Lindeman, R. E., Murphy, M. W., Agrimson, K. S., Gewiss, R. L., Bardwell, V. J., Gearhart, M. D., et al. (2021). The conserved sex regulator DMRT1 recruits SOX9 in sexual cell fate reprogramming. *Nucleic acids Res.* 49 (11), 6144–6164. doi:10.1093/nar/gkab448

Luo, D., Liu, Y., Chen, J., Xia, X., Cao, M., Cheng, B., et al. (2015). Direct production of $XY^{DMY-}$ sex reversal female medaka (*Oryzias latipes*) by embryo microinjection of TALENs. *Sci. Rep.* 5 (1), 14057–14115. doi:10.1038/srep14057

Marchand, O., Govoroun, M., D'Cotta, H., McMeel, O., Lareyre, J. J., Bernot, A., et al. (2000). DMRT1 expression during gonadal differentiation and spermatogenesis in the rainbow trout, *Oncorhynchus mykiss*. *Biochim. Biophys. Acta (BBA)-Gene Struct. Expr.* 1493 (1-2), 180–187. doi:10.1016/s0167-4781(00)00186-x

Matson, C. K., Murphy, M. W., Sarver, A. L., Griswold, M. D., Bardwell, V. J., and Zarkower, D. (2011). DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature* 476 (7358), 101–104. doi:10.1038/nature10239

McCauley, D. W. (2008). SoxE, Type II collagen, and evolution of the chondrogenic neural crest. *Zool. Sci.* 25 (10), 982–989. doi:10.2108/zsj.25.982

Ottolenghi, C., Fellous, M., Barbieri, M., and McElreavey, K. (2002). Novel paralogy relations among human chromosomes support a link between the phylogeny of doublesex-related genes and the evolution of sex determination. *Genomics* 79, 333–343. doi:10.1006/geno.2002.6711

Ottolenghi, C., Omari, S., Garcia-Ortiz, J. E., Uda, M., Crisponi, L., Forabosco, A., et al. (2005). *Foxl2* is required for commitment to ovary differentiation. *Hum. Mol. Genet.* 14 (14), 2053–2062. doi:10.1093/hmg/ddi210

Paul-Prasanth, B., Matsuda, M., Lau, E. L., Suzuki, A., Sakai, F., Kobayashi, T., et al. (2006). Knock-down of *DMY* initiates female pathway in the genetic male medaka, *Oryzias latipes*. *Biochem. Biophys. Res. Commun.* 351 (4), 815–819. doi:10.1016/j.bbrc.2006.10.095

Santerre, C., Sourdaine, P., Adeline, B., and Martinez, A. S. (2014). Cg-SoxE and Cg-β-catenin, two new potential actors of the sex-determining pathway in a hermaphrodite lophotrochozoan, the Pacific oyster *Crassostrea gigas*. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 167, 68–76. doi:10.1016/j.cbpa.2013.09.018

Sheng, Y., Chen, B., Zhang, L., Luo, M., Cheng, H., and Zhou, R. (2014). Identification of *Dmrt* genes and their up-regulation during gonad transformation in the swamp eel (*Monopterus albus*). *Mol. Biol. Rep.* 41 (3), 1237–1245. doi:10.1007/s11033-013-2968-6

Su, L., Zhou, F., Ding, Z., Gao, Z., Wen, J., Wei, W., et al. (2015). Transcriptional variants of *Dmrt1* and expression of four *Dmrt* genes in the blunt snout bream, *Megalobrama amblycephala*. *Gene* 573 (2), 205–215. doi:10.1016/j.gene.2015.07.044

Sun, H., Zhou, Z., Dong, Y ., Yang, A., Jiang, B., Gao, S., et al. (2013). Identification and expression analysis of two Toll-like receptor genes from sea cucumber (*Apostichopus japonicus*). *Fish shellfish Immunol.* 34, 147–158. doi:10.1016/j.fsi.2012.10.014

Sun, Z. H., Wei, J. L., Cui, Z. P., Han, Y. L., Chang, Y. Q., Song, J., et al. (2021). Identification and functional characterization of *piwi1* gene in sea cucumber, *Apostichopus japonicus*. *Comp. Biochem. Physiol. Part B Biochem. Mol. Biol.* 252, 110536. doi:10.1016/j.cbpb.2020.110536

Sun, J. J., Sun, Z. H., Wei, J. L., Ding, J., Song, J., and Chang, Y. Q. (2022). Identification and functional analysis of *foxl2* and *nodal* in sea cucumber, *Apostichopus japonicus*. *Gene Expr. Patterns* 44, 119245. doi:10.1016/j.gep.2022.119245

Tian, Y., Jiang, Y., Shang, Y., Zhang, Y. P., Geng, C. F., Wang, L. Q., et al. (2017). Establishment of *lysozyme* gene RNA interference systemand its involvement in salinity tolerance in sea cucumber (*Apostichopus japonicus*). *Fish Shellfish Immunol.* 65, 71–79. doi:10.1016/j.fsi.2017.03.046

Tian, W. J., He, G. D., Qin, L. J., Li, D. D., Meng, L. L., Huang, Y., et al. (2021). Genome-wide analysis of the *nramp* gene family in potato (*Solanum tuberosum*): Identification, expression analysis and response to five heavy metals stress. *Ecotoxicol. Environ. Saf.* 208, 111661. doi:10.1016/j.ecoenv.2020.111661

Vining, B., Ming, Z., Bagheri-Fam, S., and Harley, V. (2021). Diverse regulation but conserved function: SOX9 in vertebrate sex determination. *Genes* 12 (4), 486. doi:10.3390/genes12040486

Wan, H., Zhong, J., Zhang, Z., Zou, P., and Wang, Y. (2021). Discovery of the *dmrt* gene family members based on transcriptome analysis in mud crab *scylla paramamosain*. *Gene* 145576, 145576. doi:10.1016/j.gene.2021.145576

Wang, Z., Miyake, T., Edwards, S. V., and Amemiya, C. T. (2006). Tuatara (*Sphenodon*) genomics: BAC library construction, sequence survey, and application to the *dmrt* gene family. *J. Hered.* 97 (6), 541–548. doi:10.1093/jhered/esl040

Wang, F., Yu, Y., Ji, D., and Li, H. (2012). The DMRT gene family in amphioxus. *J. Biomol. Struct. Dyn.* 30 (2), 191–200. doi:10.1080/07391102.2012.677770

Wang, H., Liu, S., Cui, J., Li, C., Qiu, X., Chang, Y., et al. (2014). Characterization and expression analysis of microRNAs in the tube foot of sea cucumber Apostichopus japonicus. *PloS one* 9 (11), e111820. doi:10.1371/journal.pone.0111820

Watanabe, M., Yasuoka, Y., Mawaribuchi, S., Kuretani, A., Ito, M., Kondo, M., et al. (2017). Conservatism and variability of gene expression profiles among homeologous transcription factors in *xenopus laevis*. *Dev. Biol.* 426, 301–324. doi:10.1016/j.ydbio.2016.09.017

Wei, L., Li, X., Li, M., Tang, Y., Wei, J., and Wang, D. (2019). Dmrt1 directly regulates the transcription of the testis-biased *Sox9b* gene in Nile tilapia (*Oreochromis niloticus*). *Gene* 687, 109–115. doi:10.1016/j.gene.2018.11.016

Wei, J. L., Cong, J. J., Sun, Z. H., Song, J., Zhao, C., and Chang, Y. Q. (2021). A rapid and reliable method for genetic sex identification in sea cucumber, Apostichopus japonicus. *Aquaculture* 543, 737021. doi:10.1016/j.aquaculture.2021.737021

Wexler, J. R., Plachetzki, D. C., and Kopp, A. (2014). Pan-metazoan phylogeny of the *dmrt* gene family: A framework for functional studies. *Dev. Genes Evol.* 224 (3), 175–181. doi:10.1007/s00427-014-0473-0

Xu, X. P., Chen, X. H., Shen, X., Chen, R. Z., Zhu, C., Zhang, Z. H., et al. (2021). Genome-wide identification and characterization of *dead-box helicase* family associated with early somatic embryogenesis in *dimocarpus longan lour*. *J. Plant Physiol.* 258-259, 153364. doi:10.1016/j.jplph.2021.153364

Yamaguchi, A., Lee, K. H., Fujimoto, H., Kadomura, K., Yasumoto, S., and Matsuyama, M. (2006). Expression of the *dmrt* gene and its roles in early gonadal development of the Japanese pufferfish *Takifugu rubripes*. *Comp. Biochem. Physiol. Part D Genomics Proteomics* 1 (1), 59–68. doi:10.1016/j.cbd.2005.08.003

Yan, J., Jing, J., Mu, X., Du, H., Tian, M., Wang, S., et al. (2013). A genetic linkage map of the sea cucumber (*Apostichopus japonicus*) based on microsatellites and SNPs. *Aquaculture* 404, 1–7. doi:10.1016/j.aquaculture.2013.04.011

Yan, M., Sui, J., Sheng, W., Shao, M., and Zhang, Z. (2013). Expression pattern of *vasa* in gonads of sea cucumber *Apostichopus japonicus* during gametogenesis and reproductive cycle. *Gene Expr. Patterns* 13 (5-6), 171–176. doi:10.1016/j.gep.2013.03.001

Yang, H., Hamel, J. F., and Mercier, A. (2015). *The sea cucumber* apostichopus japonicus*: History, biology and aquaculture*. Amsterdam: Elsevier: Academic Press.

Yang, D., Zhang, Z., Liang, S., Yang, Q., Wang, Y., and Qin, Z. (2017). A novel role of Krüppel-like factor 4 in Zhikong scallop *Chlamys farreri* during spermatogenesis. *Plos one* 12 (6), e0180351. doi:10.1371/journal.pone.0180351

Yang, J., Hu, Y., Han, J., Xiao, K., Liu, X., Tan, C., et al. (2020). Genome-wide analysis of the Chinese sturgeon *sox* gene family: Identification, characterisation and expression profiles of different tissues. *J. Fish Biol.* 96, 175–184. doi:10.1111/jfb.14199

Yu, J., Zhang, L., Li, Y., Li, R., Zhang, M., Li, W., et al. (2017). Genome-wide identification and expression profiling of the SOX gene family in a bivalve mollusc Patinopecten yessoensis. *Gene* 627, 530–537. doi:10.1016/j.gene.2017.07.013

Yu, Y. Q., Ma, W. M., Zeng, Q. G., Qian, Y. Q., Yang, J. S., and Yang, W. J. (2014). Molecular cloning and sexually dimorphic expression of two *Dmrt* genes in the giant freshwater prawn, *Macrobrachium rosenbergii*. *Agric. Res.* 3 (2), 181–191. doi:10.1007/s40003-014-0106-x

Yu, Z., Zhou, Y., Yang, H., and Hu, C. (2014). Survival, growth, food availability and assimilation efficiency of the sea cucumber *Apostichopus japonicus* bottom-cultured under a fish farm in southern China. *Aquaculture* 426, 238–248. doi:10.1016/j.aquaculture.2014.02.013

Zafar, I., Rather, M. A., and Dhandare, B. C. (2019). Genome-wide identification of doublesex and Mab-3-Related transcription factor (*dmrt*) genes in nile tilapia (*Oreochromis niloticus*). *Biotechnol. Rep.* 24, e00398. doi:10.1016/j.btre.2019.e00398

Zhang, X. J., Sun, L. N., Yuan, J. B., Sun, Y. M., Yi, G., Zhang, L. B., et al. (2017). The sea cucumber genome provides insights into morphological evolution and visceral regeneration. *Plos Biol.* 15 (10), e2003790. doi:10.1371/journal.pbio.2003790

Zhang, J., Han, X., Wang, J., Liu, B. Z., Wei, J. L., Zhang, W. J., et al. (2019). Molecular cloning and sexually dimorphic expression analysis of *nanos2* in the sea urchin, *Mesocentrotus nudus*. *Int. J. Mol. Sci.* 20 (11), 2705. doi:10.3390/ijms20112705

Zhang, H., Tan, K., Li, S., Ma, H., and Zheng, H. (2020). Genome-wide analysis of *traf* gene family and its response to bacterial infection in noble scallop *chlamys nobilis* with different carotenoids content. *Aquaculture* 535 (5625), 736309. doi:10.1016/j.aquaculture.2020.736309

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 09 February 2023
DOI 10.3389/fgene.2023.1077814



OPEN ACCESS

EDITED BY
Siti Nor,
University of Malaysia Terengganu, Malaysia

REVIEWED BY
Hui Qiao,
Freshwater Fisheries Research Center (CAFS), China
Zhongkai Wang,
Qingdao Agricultural University, China

*CORRESPONDENCE
Weiji Wang,
wangwj@ysfri.ac.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

RECEIVED 23 October 2022
ACCEPTED 27 January 2023
PUBLISHED 09 February 2023

CITATION
Lyu D, Sun S, Shan X and Wang W (2023), Inbreeding evaluation using microsatellite confirmed inbreeding depression in growth in the *Fenneropenaeus chinensis* natural population.
*Front. Genet.* 14:1077814.
doi: 10.3389/fgene.2023.1077814



# Inbreeding evaluation using microsatellite confirmed inbreeding depression in growth in the *Fenneropenaeus chinensis* natural population

Ding Lyu[1,2†], Song Sun[1,2†], Xiujuan Shan[1,2] and Weiji Wang[1,2*]

[1]Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China, [2]Function Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology, Qingdao, China

Understanding inbreeding depressions (IBDs), the effect on the phenotypic performance of inbreeding, is of major importance for evolution and conservation genetics. Inbreeding depressions in aquatic animals were well documented in a domestic or captive population, while there is less evidence of inbreeding depression in natural populations. Chinese shrimp, *Fenneropenaeus chinensis*, is an important species in both aquaculture and fishery activities in China. To investigate inbreeding depression in natural populations, four *Fenneropenaeus chinensis* natural populations (Huanghua, Qinhuangdao, Qingdao, and Haiyang) were collected from the Bohai and Yellow seas. Microsatellite markers were used to evaluate individual inbreeding coefficients (F) of all samples. Furthermore, the effects of inbreeding on growth traits were investigated. The results showed marker-based F was continuous and ranged from 0 to 0.585, with an average of 0.191 ± 0.127, and there was no significant difference among the average F of the four populations. Regression analysis using the four populations showed inbreeding had a very significant ($p < 0.01$) effect on body weight. When analyzing a single population, regression coefficients were also all negative and those in Huanghua and in Qingdao were significant at the level of $p < 0.05$ and $< 0.01$, respectively. Inbreeding depressions, expressed as the percent change in body weight per 10% increase in F, were 2.75% in Huanghua, 2.22% in Qingdao, and 3.69% in all samples. This study provided a piece of rare evidence of inbreeding depression in natural populations and also guidance toward the conservation of wild *Fenneropenaeus chinensis* resources.

KEYWORDS

**inbreeding depression, *Fenneropenaeus chinensis*, microsatellite, body weight, inbreeding coefficient**

## 1 Introduction

Inbreeding is defined as the mating of individuals that are related by ancestry and results in the reduction of heterozygosity (Falconer and Mackay, 1996). The inbreeding coefficient (F) is a measure of an inbreeding level and can be defined as both the probability that two alleles at any given locus are identical by descent (alleles are descendants from a single ancestor) and the probable proportion of an individual's loci containing genes that are identical by descent (Falconer and Mackay, 1996; Bourdon, 1997). Inbreeding depression (IBD) is the effect of

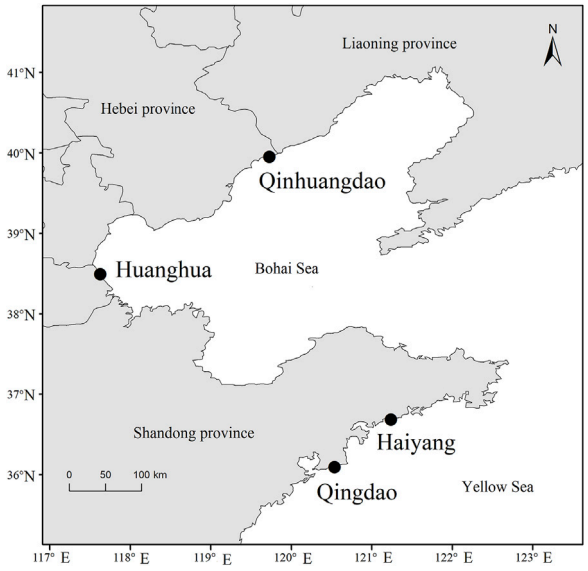



FIGURE 1
Distribution of *Fenneropenaeus chinensis* sampling locations.

**TABLE 1** Sampling locations, time, and number of the four populations.

| Location | Sea area | Sampling time | Number |
|---|---|---|---|
| Huanghua | Bohai Sea | 11 September 2021 | 194 |
| Qinhuangdao | Bohai Sea | 17 September 2021 | 136 |
| Qingdao | Yellow Sea | 1 to 13 April, 2022 | 159 |
| Haiyang | Yellow Sea | 11 to 19 April, 2022 | 75 |

inbreeding, measured as the reduction in mean phenotypic performance with increasing levels of inbreeding within a population (Falconer and Mackay, 1996; Lynch and Walsh, 1998). The existence of IBD has long been known, especially for fitness traits. Understanding IBD is of major importance for the evolution and conservation of genetics. These effects have been well documented in livestock species (reviewed by Leroy, 2014), and also in aquatic animals (Keys et al., 2004; Zheng et al., 2012; Luo et al., 2014; Gao et al., 2015). However, most of these study examples were carried out in the domestic or captive population, and relatively less evidence of IBD was illustrated in natural populations (Hoffman et al., 2014). The main reason is that F was very accessible in a captive population with pedigree information, in which F of any individual can be obtained by calculating its parents' coancestry (Falconer and Mackay, 1996). However, investigating the inbreeding level and IBD in natural populations can also be of great significance, and IBD is one of the core research fields in conservation genetics (Frankham et al., 2009). An alternative approach to the pedigree method is calculating F based on molecular markers such as microsatellites (Ritland, 1996; Lynch and Ritland, 1999; Milligan, 2003; Wang, 2007). Compared to the traditional method, F based on a molecular marker can be obtained directly, without the need for pedigree information. As a result, it provided the possibility to study IBD in natural populations.

Chinese shrimp, *F. chinensis*, is an important species in both aquaculture and fishery activities in China. Over the past few decades, the natural population of *F. chinensis* was largely reliant on released shrimps to maintain its size, and their contribution to the total landings has been consistently > 90% (Wang et al., 2006). There is a general belief that a genetic threat of the loss of variation in wild populations was one of the main concerns about an artificial propagation release (Aho et al., 2006; Araki and Schmid, 2010). Also, it was inferred that continuous artificial propagation and release had lowered the level of genetic diversity of *F. chinensis* in Chinese stocks (Wang et al., 2006). Although a previous study demonstrated that inbreeding has a negative effect on economic traits (especially on growth) in the *F. chinensis* breeding population by comparing different levels of inbreeding (Luo et al., 2014), this phenomenon has not been demonstrated under natural conditions. Such a study is important both because inbreeding may affect the extinction risk in wild populations and because understanding IBD is of major importance for the conservation genetics of *F. chinensis*.

In the current study, multiple *F. chinensis* natural populations were collected from the Bohai and Yellow seas in northern China. Microsatellite markers were used to analyze these *F. chinensis* samples to calculate individual F. Furthermore, its effects on growth were investigated. The results obtained in this study should provide evidence of IBD in natural populations and also guidance to the conservation of *F. chinensis* resources.

# 2 Materials and methods

## 2.1 Experimental materials

The *F. chinensis* samples were collected from four locations in the Bohai and Yellow seas: Huanghua and Qinhuangdao populations in the Bohai Sea and Qingdao and Haiyang populations in the Yellow Sea (Figure 1). The sampling time in Bohai Sea was in the autumn of 2021 and that in the Yellow Sea was in the spring of 2022. The sample size in each population ranged from 75–194, and the total was 564 (Table 1). All the samples in Qingdao and Haiyang were females because males died after mating in October or November. Those in Huanghua and Qinhuangdao contained both sexes. The body weight of each individual was measured and all samples were transported to the laboratory in liquid nitrogen and stored at −80 °C until analysis.

Genomic DNA was extracted from swimming legs in all individuals using standard phenol-chloroform procedures (Sambrook et al., 1989). The primer sequence, fluorescent dyes and anneal of genotyping microsatellite are shown in Table 2. PCR thermal cycling was performed as follows: an initial denaturing at 94°C for 5 min, followed by 30 cycles including 30 s denaturing at 94°C, 30 s annealing at locus-specific temperatures, and 40 s extension at 72°C, and then with a final extension at 72°C for 5 min. The PCR products were separated by an ABI-3130 automated Genetic Analyzer (Applied Biosystems). Alleles from the microsatellite loci were sized with a GeneScanTM-500 LIZ Size Standard (Applied Biosystems) and scored using GeneMapper™ V4.1 (Applied Biosystems).

## 2.2 Inbreeding coefficient calculation

The genetic diversity parameters, including the number of alleles (N), observed heterozygosity (Ho), expected heterozygosity (He), and polymorphism information content (PIC) at 11 genotyping

**TABLE 2** Primer sequences, fluorescent dyes, and anneal of microsatellites.

| Microsatellite | Primer sequence (5′–3′) | Fluorescent dye | Anneal (°C) |
|---|---|---|---|
| EN0033 | F: CCTTGACACGGCATTGATTGG | 6-FAM | 64 |
| | R: TACGTTGTGCAAACGCCAAGC | | |
| RS0622 | F: CAGTCCGTAGTTCATACTTGG | HEX | 66 |
| | R: ACATGCCTTTGTGTGAAAACG | | |
| RS1101 | F: CGAGTGGCAGCGAGTCCT | ROX | 52 |
| | R: TATTCCCACGCTCTTGTC | | |
| RS0683 | F: CACTCACTTATGTCACACTGC | TAMRA | 66 |
| | R: ACACACCAACACTCAATCTCC | | |
| EN0113 | F: TGTCAAGAGAGCGAGAGGGAGG | 6-FAM | 65 |
| | R: TGTCAAGAGAGCGAGAGGGAGG | | |
| BM29561 | F: AACAGACCACATACGGGAC | HEX | 58 |
| | R: TTTTCGGAAGTAACATCACA | | |
| RS0916 | F: GGCTAATGATAATAATGCTG | ROX | 56 |
| | R: CGTTGTTGTTGCTGTTG | | |
| RS0779 | F: ATGACACTCAAATCAAAG | TAMRA | 50 |
| | R: CAGAATAACATCATTACTAC | | |
| FCKR009 | F: GCACGAAAACACATTAGTAGGA | 6-FAM | 52 |
| | R: ATATCTGGAATGGCAAAGAGTC | | |
| FCKR002 | F: CTCAACCCTCACCTCAGGAACA | HEX | 60 |
| | R: AATTGTGGAGGCGACTAAGTTC | | |
| FCKR0013 | F: GCACATATAAGCACAAACGCTC | ROX | 61 |
| | R: CTCTCTCGCAATCTCTCCAACT | | |

microsatellite loci were obtained using CERVUS software (Kalinowski et al., 2007). A fixation index ($F_{st}$) was used to measure genetic differentiation among populations and calculated using Arlequin software (Excoffier et al., 2005). There are multiple ways of calculating F from microsatellite genotyping data (Ritland, 1996; Lynch and Ritland, 1999; Milligan, 2003; Wang, 2007), and a triadic likelihood method (Wang, 2007) was adopted in the current study. This was because it was also proven to be the most accurate method to calculate the paired coancestry, having either the lowest root mean error or close to the smallest one in different population structures and sizes, number of loci, and alleles (Wang, 2007; Hammerly et al., 2013). The triadic likelihood method was carried out in coancestry software (Wang, 2011) to obtain F from microsatellite genotyping data.

## 2.3 Inbreeding effect evaluation on body weight

A linear-regression model was used to evaluate the effects of F on body weight:

$$y = a + b \times F + e, \quad (1)$$

where $y$ is the measured value for body weight; $a$ is the $y$-intercept; $b$ is the regression coefficient of F; and $e$ is the residual term. The residual term in this model includes individual genetic effects, day-age effects, and error terms, all of which are unrelated to inbreeding. In particular, since the interaction effect between sex and F was not significant ($p > 0.05$), the analysis of sex effects in Huanghua and Qinhuangdao populations were also subsumed into the residual term.

*IBD*, expressed as the percent change in phenotype per 10% increase in F, was calculated with the following equation:

$$IBD = \left[ (b \times 0.1)/a \right] \times 100, \quad (2)$$

where $a$ and $b$ are $y$-intercept from the regression of $y$ on F and the regression coefficient of F, respectively (see Eq. 1). All of the aforementioned statistical analysis processes were performed with corresponding functions in R software (R Development Core Team, 2013).

# 3 Results

## 3.1 Inbreeding coefficient calculation

The N, Ho, He, and PIC at each microsatellite locus are shown in Table 3, and the average $N$ was 25.18 ± 14.68. The $F_{st}$ ranged from

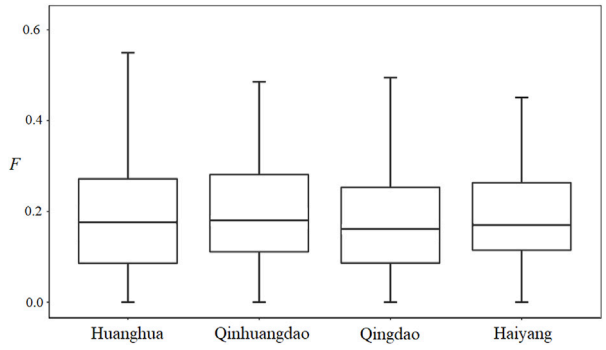



FIGURE 2
Individual F distribution of *Fenneropenaeus chinensis* in the four populations.

**TABLE 3 Number of alleles (N), observed heterozygosity (Ho), expected heterozygosity (He), and polymorphism Information content (PIC) at 11 microsatellites.**

| Microsatellite | N | Ho | He | PIC |
|---|---|---|---|---|
| EN033 | 55 | 0.728 | 0.969 | 0.966 |
| RS0622 | 35 | 0.923 | 0.954 | 0.950 |
| RS1101 | 14 | 0.762 | 0.797 | 0.767 |
| RS0683 | 38 | 0.769 | 0.927 | 0.920 |
| EN0113 | 12 | 0.782 | 0.862 | 0.845 |
| BM29561 | 31 | 0.898 | 0.898 | 0.888 |
| RS0916 | 4 | 0.485 | 0.582 | 0.490 |
| RS0779 | 11 | 0.692 | 0.872 | 0.856 |
| FCKR009 | 28 | 0.745 | 0.923 | 0.916 |
| FCKR002 | 27 | 0.825 | 0.944 | 0.939 |
| FCKR013 | 22 | 0.821 | 0.920 | 0.912 |

0.004–0.012 (Table 4), with an average of 0.008 ± 0.003, and only that between Qinhuangdao and Qingdao was significant ($p < 0.05$). The marker-based F was continuous and ranged from 0 to 0.585, with an average of 0.191 ± 0.127. The average F of Qingdao and Haiyang were 0.180 and 0.185, respectively, lower than those of Qinhuangdao and Huanghua, 0.202 and 0.194, respectively (Figure 2). However, there was no significant difference among the average F of the four populations.

## 3.2 Inbreeding effects on body weight

Relationships between F and measured body weight in each population and all four populations are shown in Figures 3, 4, respectively. Regression analysis' results are shown in Table 5. Regression analysis using four populations showed that inbreeding had a very significant ($p < 0.01$) effect on body weight. When analyzing a single population, regression coefficients were also all negative and those in Huanghua and Qingdao were significant at the level of $p < 0.05$ and $< 0.01$, respectively. However, those in Qinhuangdao and Haiyang were not significantly different from zero due to their high standard errors. In particular, when regression analysis was based on four populations, the interaction effect between the population and F was not significant. For analyses with significant regression coefficients, estimates of IBD on body weight were further calculated. Those in Huanghua and Qingdao were 2.75% and 2.22%, respectively, and when all individuals were fitted, the estimate of IBD was 3.39% (Table 5).

# 4 Discussion

We estimated individual F within four *F. chinensis* populations directly, with no requirement of pedigree records or pedigree reconstruction. This method provided an alternative approach to detecting the inbreeding level of *F. chinensis* natural populations. The individual F estimate results showed that there was no significant difference in average F among different populations. All the genetic differentiation between different populations was very low, and only that between Qinhuangdao and Qingdao was significant ($p < 0.05$). This suggested that there was no or very little genetic differentiation among these populations, which was consistent with the expected average F. Most previous studies thought of *F. chinensis* distributed along northern China as one population and suggested a lack of genetic structure in this species (Hwang et al., 1997; Quan et al., 2001; Cui et al., 2007). Furthermore, there was no difference in average F before (Huanghua and Qinhuangdao) and after (Qingdao and Haiyang) the overwintering migration, which suggested that there was no significant relationship between F and the survival rate during overwintering migration. However, the average individual F obtained in this study was consistently high. There are several possible reasons responsible for such a high F. The first is that successive years of artificial propagation and release may have reduced the effective population size of *F. chinensis* and increased the possibility of inbreeding; another possible reason is the occurrence of null alleles that might have also led to Ho in most loci to be lower than that of He.

Based on individual F calculated from microsatellites, we provided a rare piece of evidence of IBD in a natural aquatic animal population

**TABLE 4 Fixation index ($F_{st}$) among geographic populations.**

| | Huanghua | Qinhuangdao | Qingdao | Haiyang |
|---|---|---|---|---|
| Huanghua | — | 0.141 | 0.039 | 0.198 |
| Qinhuangdao | 0.007 | — | 0.093 | 0.115 |
| Qingdao | 0.012* | 0.009 | — | 0.143 |
| Haiyang | 0.004 | 0.008 | 0.007 | — |

The lower triangular elements are the fixation index ($F_{st}$) values, and the upper triangular elements are their *p*-values. The asterisk (*) indicates $F_{st}$ is significantly different from zero ($p < 0.05$).

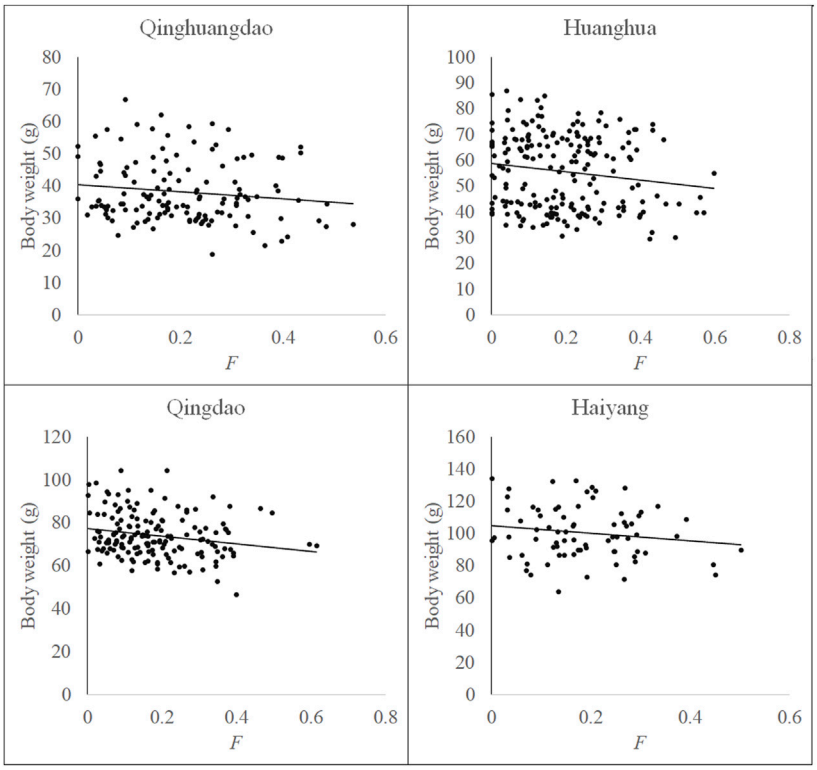



FIGURE 3
Relationship between F and body weight in each population.

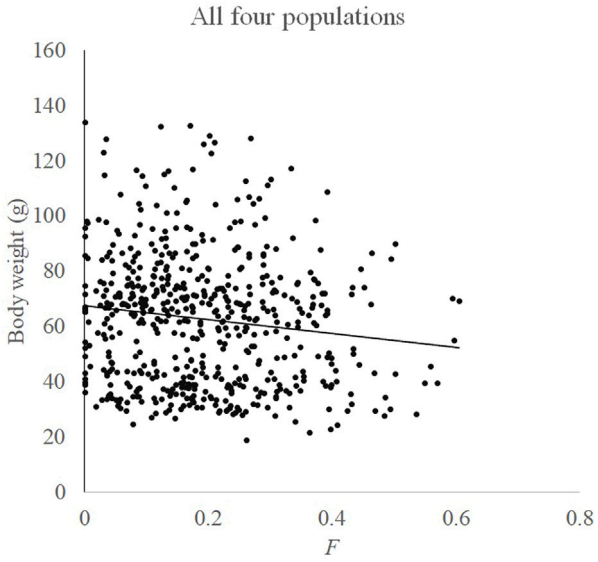



FIGURE 4
Relationship between F and body weight in all four populations.

as opposed to that in a captive population. The results of the current study showed that F had a significant negative effect on body weight in *F. chinensis* in the wild ($p < 0.01$). In addition, when the samples were divided into four populations according to where they were collected, regression analysis results in two of them supporting this conclusion. Insignificant regression coefficients in another two can likely be attributed to sampling errors which are associated with both a small sample size and a small number of microsatellite loci. Moreover, the interaction effect between the population and F was not significant, which also suggested that IBD should be universal in different populations. Similar to microsatellites, single nucleotide polymorphism (SNP) is another available molecular marker to estimate individual F (Ritland, 1996; Lynch and Ritland, 1999), and F based on whole-genome high-density SNPs should obviously be more accurate. However, the high cost restricts its routine usage in inbreeding level monitoring. If there are no applicable SNP data, using a small number of microsatellites is also an alternative method because microsatellite can provide more marker information contents due to its high polymorphism (Wang, 2016).

Although IBD typically was detected in fitness-related traits (e.g., survival and reproductive traits) (Falconer and Mackay, 1996; Lynch and Walsh, 1998), there was also numerous research on IBD of growth traits in aquaculture species. In another study on *F. chinensis*, Luo et al. (2014) estimated the IBD of body weight as 4.16%–4.74%, which was slightly larger than those in the current study. However, it should be noted that the estimated IBD in Luo et al. (2014) was obtained by comparing several designed high inbreeding levels ($F$ = 0.25, 0.375, and 0.50) in a breeding population. Some research has indicated fast inbreeding (mating between close relatives, such as full sibs) caused more harm than slow inbreeding (Day et al., 2003; Thodesen et al., 2005). This view might explain the difference in results between these two studies. In addition, the occurrence of null alleles may affect the accuracy of the estimated F, and the IBD scales were also affected to

**TABLE 5** Results from regressions of body weight on F and IBD expressed as percent change in phenotype per 10% increase in F.

| Population | *a* | *b*±se | *p* | IBD (%) |
|---|---|---|---|---|
| Huanghua | 58.80 | −16.18 ± 7.67 | 0.036* | −2.75 |
| Qinhuangdao | 40.36 | −10.86 ± 6.85 | 0.116 | — |
| Qingdao | 77.31 | −17.31 ± 6.58 | 0.009** | −2.22 |
| Haiyang | 104.73 | −23.70 ± 17.36 | 0.176 | — |
| All | 67.35 | −24.70 ± 7.86 | 0.002** | −3.69 |

The asterisk (*) indicates *b* is significantly different from zero ($p < 0.05$), and the double asterisk (**) indicates *b* is very significantly different from zero ($p < 0.01$).

some extent. Thodesen et al. (2005) summarized previous research on IBD in aquaculture species and found the average IBD of slow inbreeding on growth traits ranging from 0% to 13% with an average of 2.4%, and the results in the current study were fairly consistent with previous research.

In this study, we demonstrated the phenomenon of IBD in a natural population of *F. chinensis* for the first time (to the best of our knowledge). Although a study has demonstrated the existence of IBD in *F. chinensis* breeding population (Luo et al., 2014), the finding in the current study is still of great significance. This is because IBD under laboratory and captive conditions may not be representative of that under natural conditions (Crnokrak and Roff, 1999), and IBD may vary with environmental conditions (Armbruster and Reed, 2005). Over the past few decades, the natural population of *F. chinensis* in China was largely reliant on released shrimps to maintain its size (Deng and Zhuang, 2000; Wang et al., 2006). Due to its strong reproduction ability, often only a few gravid females are needed to produce enough released offspring for a hatchery. Some scholars have long been concerned about the effective population size decreasing in the *F. chinensis* natural population (Wang et al., 2006; Song et al., 2019). Therefore, to avoid IBD, it is very important to maintain the effective population size of the releasing population for *F. chinensis*.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

## Author contributions

LD wrote the main manuscript text and prepared all figures and tables; LD and SS analyzed the experimental data; and SX and WW guided the research. All authors reviewed the manuscript.

## Funding

This study was financially supported by the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No. 2021QNLM050103-3) and the Central Public-Interest Scientific Institution Basal Research Fund, YSFRI, CAFS (No. 20603022021002).

## Conflict of interest

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

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Aho, T., Rönn, J., Piironen, J., and Björklund, M. (2006). Impacts of effective population size on genetic diversity in hatchery reared Brown trout (Salmo trutta L.) populations ective population size on genetic diversity in hatchery reared Brown trout (*Salmo trutta* L.) populations. *Aquaculture* 253, 244–248. doi:10.1016/j.aquaculture.2005.09.013

Araki, H., and Schmid, C. (2010). Is hatchery stocking a help or harm?: Evidence, limitations and future directions in ecological and genetic surveys. *Aquaculture* 308, S2–S11. doi:10.1016/j.aquaculture.2010.05.036

Armbruster, P., and Reed, D. H. (2005). Inbreeding depression in benign and stressful environments. *Heredity* 95 (3), 235–242. doi:10.1038/sj.hdy.6800721

Bourdon, R. M. (1997). *Understanding animal breeding*. 2nd ed. New Jersey, USA: Prentice-Hall.

Crnokrak, P., and Roff, D. A. (1999). Inbreeding depression in the wild. *Heredity* 83 (3), 260–270. doi:10.1038/sj.hdy.6885530

Cui, Z., Li, C., Jang, I., and Chu, K. (2007). Lack of genetic differentiation in the shrimp *Penaeus chinensis* in the Northwestern Pacific. *Biochem. Genet.* 45, 579–588.

Day, S. B., Bryant, E. H., and Meffert, L. M. (2003). The influence of variable rates of inbreeding on fitness, environmental responsiveness, and evolutionary potential. *Evolution* 57 (6), 1314–1324. doi:10.1111/j.0014-3820.2003.tb00339.x

Deng, J. Y., and Zhuang, Z. M. (2000). The cause of recruitment variation of *Penaeus chinensis* in the Bohai Sea. *J. Fish. Sci. China.* 7, 125–128.

Excoffier, L., Laval, L. G., and Schneider, S. (2005). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform.* 1, 117693430500100–117693430500150. doi:10.1177/117693430500100003

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative genetics*. 4th ed. Harlow, Essex, England: Longman.

Frankham, R., Ballou, J. D., and Briscoe, D. A. (2009). *Introduction to conservation genetics*. 2nd edn. Cambridge, UK: Cambridge University Press.

Gao, B., Liu, P., Li, J., Wang, Q., and Han, Z. (2015). Effect of inbreeding on growth and genetic diversity of *Portunus trituberculatus* based on the full-sibling inbreeding families. *Aquac. Int.* 23, 1401–1410. doi:10.1007/s10499-015-9892-9

Hammerly, S. C., Morrow, M. E., and Johnson, J. A. (2013). A comparison of pedigree- and DNA-based measures for identifying inbreeding depression in the critically endangered Attwater's Prairie-chicken. *Mol. Ecol.* 22, 5313–5328. doi:10.1111/mec.12482

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A. S., et al. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proc. Natl. Acad. Sci.* 111 (10), 3775–3780. doi:10.1073/pnas.1318945111

Hwang, G. L., Lee, Y. C., and Chang, C. S. (1997). Mitochondrial DNA analysis of the fleshy prawn (*Penaeus chinensis*) for stock discrimination in the Yellow Sea. *J. Korean Fish. Soc.* 30, 88–94.

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16 (5), 1099–1106. doi:10.1111/j.1365-294X.2007.03089.x

Keys, S. J., Crocos, P. J., Burridge, C. Y., Coman, G. J., Davis, G. P., and Preston, N. P. (2004). Comparative growth and survival of inbred and outbred *Penaeus* (*marsupenaeus) japonicus*, reared under controlled environment conditions: Indications of inbreeding depression. *Aquaculture* 241, 151–168. doi:10.1016/j.aquaculture.2004.08.039

Leroy, G. (2014). Inbreeding depression in livestock species: Review and meta-analysis. *Anim. Genet.* 45 (5), 618–628. doi:10.1111/age.12178

Luo, K., Kong, J., Luan, S., Meng, X. H., Zhang, T. S., and Wang, Q. Y. (2014). Effect of inbreeding on survival, WSSV tolerance and growth at the postlarval stage of experimental full-sibling inbred populations of the Chinese shrimp *Fenneropenaeus chinensis*. *Aquaculture* 420, 32–37. doi:10.1016/j.aquaculture.2013.10.030

Lynch, M., and Ritland, K. (1999). Estimation of pairwise relatedness with molecular markers. *Genetics* 152, 1753–1766. doi:10.1093/genetics/152.4.1753

Lynch, M., and Walsh, B. (1998). *Genetics and analysis of Quantitative traits*. Sunderland, MA: Sinauer Associates, 360–361.

Milligan, B. G. (2003). Maximum-likelihood estimation of relatedness. *Genetics* 163, 1153–1167. doi:10.1093/genetics/163.3.1153

Quan, J. X., Lu, X. M., Zhuang, Z. M., Dai, J. X., Deng, J. Y., and Zhang, Y. P. (2001). Low genetic variation of *Penaeus chinensis* as revealed by mitochondrial COI and 16S rRNA gene sequences. *Biochem. Genet.* 39, 279–284. doi:10.1023/a:1010234816514

R Development Core Team (2013). *R: A Language and environment for statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ritland, K. (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. *Genet. Res.* 67, 175–185. doi:10.1017/s0016672300033620

Sambrook, J., Fristch, E. E., and Maniatis, T. (1989). *Molecular Cloning: A laboratory manual*. New York: Cold Spring Harbor Laboratory.

Song, N., Li, P., Zhang, X., and Gao, T. (2019). Genetic diversity status and conservation priority of hatchery-produced offspring populations of *Fenneropenaeus chinensis* by microsatellite DNA. *Mar. Freshw. Res.* 71 (12), 1592–1601. doi:10.1071/mf19235

Thodesen, J., Hu, H. L., and Kong, J. (2005). Inbreeding and its impact on aquaculture. *J. Fish. China* 29 (6), 849–856.

Wang, J. (2011). Coancestry: A program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Mol. Ecol. Resour.* 11, 141–145. doi:10.1111/j.1755-0998.2010.02885.x

Wang, J. (2016). Pedigrees or markers: Which are better in estimating relatedness and inbreeding coefficient? *Theor. Popul. Biol.* 107, 4–13. doi:10.1016/j.tpb.2015.08.006

Wang, J. (2007). Triadic IBD coefficients and applications to estimating pairwise relatedness. *Genet. Res.* 89, 135–153. doi:10.1017/S0016672307008798

Wang, Q., Zhuang, Z., Deng, J., and Ye, Y. (2006). Stock enhancement and translocation of the shrimp *Penaeus chinensis* in China. *Fish. Res.* 80 (1), 67–79. doi:10.1016/j.fishres.2006.03.015

Zheng, H., Li, L., and Zhang, G. (2012). Inbreeding depression for fitness-related traits and purging the genetic load in the hermaphroditic bay scallop *Argopecten irradians irradians* (*Mollusca: Bivalvia*). *Aquaculture* 366-367, 27–33. doi:10.1016/j.aquaculture.2012.08.029

# Frontiers in Genetics

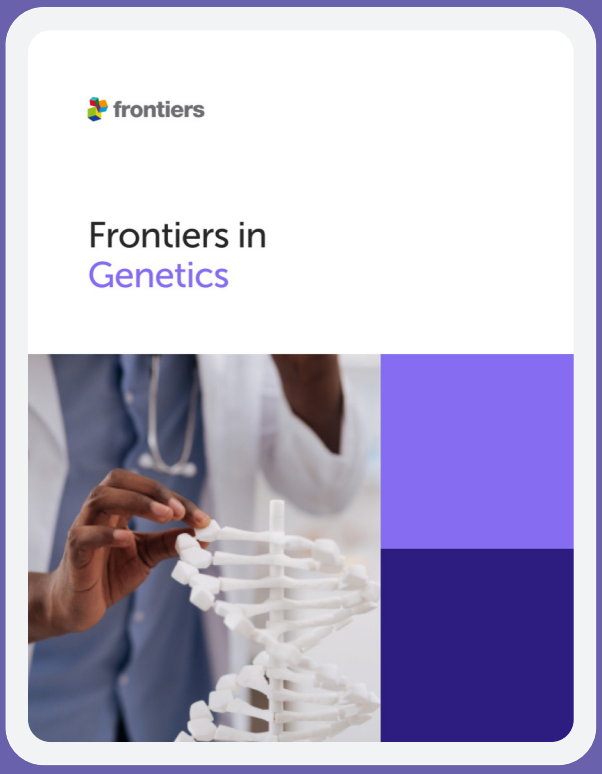
**Highlights genetic and genomic inquiry relating to all domains of life**

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

---

## Discover the latest Research Topics

See more →




**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact