GENETICS AND MOLECULAR BREEDING IN AQUACULTURE ANIMALS

EDITED BY: Zexia Gao, Alexandre Wagner Silva Hilsdorf, Qiang Lin and Li Zhou

PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-854-1
DOI 10.3389/978-2-83250-854-1

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

GENETICS AND MOLECULAR BREEDING IN AQUACULTURE ANIMALS

Topic Editors:
Zexia Gao, Huazhong Agricultural University, China
Alexandre Wagner Silva Hilsdorf, University of Mogi das Cruzes, Brazil
Qiang Lin, South China Sea Institute of Oceanology, Chinese Academy of Sciences (CAS), China
Li Zhou, Institute of Hydrobiology, Chinese Academy of Sciences (CAS), China

Citation: Gao, Z., Hilsdorf, A. W. S., Lin, Q., Zhou, L., eds. (2022). Genetics and Molecular Breeding in Aquaculture Animals. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-854-1

Table of Contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 10 November 2022
DOI 10.3389/fgene.2022.1071303

OPEN ACCESS

EDITED AND REVIEWED BY
Martino Cassandro,
University of Padua, Italy

*CORRESPONDENCE
Ze-Xia Gao,
gaozx@mail.hzau.edu.cn

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

RECEIVED 16 October 2022
ACCEPTED 31 October 2022
PUBLISHED 10 November 2022

CITATION
Yu Y, Hilsdorf AWS, Zhou L, Lin Q and Gao Z-X (2022), Editorial: Genetics and molecular breeding in aquaculture animals.
Front. Genet. 13:1071303.
doi: 10.3389/fgene.2022.1071303

Editorial: Genetics and molecular breeding in aquaculture animals

Yue Yu[1], Alexandre Wagner Silva Hilsdorf[2], Li Zhou[3], Qiang Lin[4] and Ze-Xia Gao[1]*

[1]Key Lab of Freshwater Animal Breeding, Ministry of Agriculture and Rural Affairs/Key Lab of Agricultural Animal Genetics, College of Fisheries, Breeding and Reproduction of Ministry of Education/Engineering Research Center of Green Development for Conventional Aquatic Biological Industry in the Yangtze River Economic Belt of Ministry of Education/Engineering Technology Research Center for Fish Breeding and Culture in Hubei Province, Huazhong Agricultural University, Wuhan, China, [2]Unit of Biotechnology, University of Mogi das Cruzes, Mogi das Cruzes, Brazil, [3]State Key Laboratory of Freshwater Ecology and Biotechnology, Hubei Hongshan Laboratory, The Innovation Academy of Seed Design, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China, [4]CAS Key Laboratory of Tropical Marine Bio-Resources and Ecology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

KEYWORDS

aquaculture, genetics, genomics, molecular breeding, SNP, genome editing

Editorial on the Research Topic
Genetics and Molecular Breeding in Aquaculture Animals

As a major source of "blue food" (Gephart et al., 2021), aquaculture is an important part of meeting the growing human demand for protein, and is also the fastest growing field of food production in the world (FAO, 2020). Selective breeding offers enormous potential to improve important commercial characters and increase commercial efficiency in aquaculture by providing cumulative and permanent genetic improvement of farmed species. Compared to traditional genetic improvement strategies, molecular marker-assisted selection (MAS) rapidly screens for dominant formations within a population by locating genetic markers that are closely linked to the target trait, which greatly saves time and economic costs to increase breeding efficiency (Collard and Mackill, 2008). During the past few decades, better molecular markers, sequencing techniques and breeding algorithms mean that the mapping of specific genetic loci for economic characters is more efficient and precise. In particular, the advent of high-throughput sequencing technologies has greatly facilitated the rapid development of genomics and molecular biology to analyze the biological and genetic background of quantitative traits, so as to promote the application of aquaculture molecular breeding. The aim of this Research Topic is to collect latest molecular breeding model and/or high-quality research finding on aquaculture genetics and genetic breeding. At present, the Research Topic has collected twelve articles, which are groups in seven different themes (Table 1), and are summarized in detail as follows:

Theme 1 (Genetic diversity): The estimation of population genetic diversity can not only provide basic information for resource assessment and management, but lay the foundation for germplasm improvement. Wang et al. used GBS

TABLE 1 Main research themes included in the Research Topic.

No.	Theme	Species	Main findings	References
1	Genetic diversity	sea urchin *Mesocentrotus nudus*	51,738 biallelic SNPs	Wang et al.
2	Discovery of associated molecular markers and candidate genes for economic traits	oriental river prawn *Macrobrachium nipponense*	$NF_kB\alpha$ had a positive regulatory effect on testicular development	Jin et al.
		swimming crab *Portunus trituberculatus*	8 molecular markers for salt-alkali tolerance	Zhang et al.
		Pacific white shrimp *Litopenaeus Vannamei*	489 DEGs	Zhang et al.
3	Linkage maps and QTL analysis	Red swamp crayfish *Procambarus clarkii*	4,878 SNPs and 94 LGs	Guo et al.
		common carp *Cyprinus carpio*	28,416 SNPs and 50 LGs	Zhang et al.
		large yellow croaker *Larimichthys crocea*	9,885 SNPs and 24 LGs	Yu et al.
4	References genome assembly	pedunculate barnacle *Capitulum mitella*	Frist draft genome assembly	Chen et al.
5	Exploratory improvement methods for molecular breeding	Atlantic salmon *Salmo salar*	A cost-saving approach to genotyping	Dagnachew et al.
6	Quantitative genetic basis and genomic architecture of commercial traits	Atlantic salmon *S. salar*	Genetic variation and genetic correlation for smoltification traits	Khaw et al.
		Nile tilapia *Oreochromis niloticus*	Population genetic variation and QTL localization for feed efficiency traits	Barría et al.
7	Functional gene impact on omics	zebrafish *Danio rerio*	22 DEmiRNAs	Domingues et al.

(genotyping-by-sequencing) technology to genotype 80 *Mesocentrotus nudus* individuals from five populations along the China's coast with single nucleotide polymorphisms (SNPs) at the genome-wide level, and systematically described the genetic diversity and population structure of *M. nudus*. The comprehensive results of principal component analysis, admixture and phylogenetic analysis revealed a low genetic differentiation and high genetic connectivity among the five populations, suggesting that timely conservation measures should be taken for wild *M. nudus* resources.

Theme 2 (Discovery of associated molecular markers and candidate genes for economic traits): Accurate targeting of genetic markers associated with target traits is a crucial prerequisite for molecular breeding. Jin et al. confirmed the inhibitory effect of eyestalk on male development in *Macrobrachium nipponense* by histological observation of gonads. Transcriptome differential analysis of normal prawns (CG), single-side eyestalk ablation prawns (SS), and double-side eyestalk ablation prawns (DS) revealed 1,039, 1,226, and 3,682 differentially expressed genes (DEGs) between CG and SS, SS and DS, and CG and DS, respectively. $NF_kB\alpha$ was screened for further functional analysis with qPCR, RNAi and histological observation, as it was the most upregulated gene expressed in DS group. The results indicated that $NF_kB\alpha$ had a positive regulatory effect on testicular development in *M. nipponense*, and there was a positive regulatory relationship between $NF_kB\alpha$ and insulin-like androgen hormone (IAG), which confirmed the important function of $NF_kB\alpha$ in male sexual development of crustaceans for the first time. In the second report, Zhang et al. tentatively explored the salt-alkali tolerance molecular markers in *Portunus trituberculatus* by the bulked segregant analysis (BSA) strategy. Extreme populations of salt-alkali tolerance or intolerance were distinguished to construct two DNA mixing pool libraries. After comparison and association analysis, eight molecular markers that showed significantly related to phenotype were selected, including five SNPs and three indels. For another global crustacean species, *Litopenaeus Vannamei*, Zhang et al. conducted a comparative transcriptome approach to uncover the genetic mechanisms of resistance to *Vibrio parhaemolyticus* (VP_{AHPND}), one of the major pathogens of acute hepatopancreatic necrosis disease in *L. vannamei*. After the VP_{AHPND} challenge test, three resistant families and three susceptible families were picked out from 95 families for transcriptome sequencing to construct gene expression profiles. Differential gene expression analysis screened 489 DEGs, of which 19 genes were successfully verified in the offspring lines, underlining the accuracy of these markers.

Theme 3 (Linkage maps and QTL analysis): Genetic linkage mapping and QTL localization are important tools for marker-assisted selection for breeding because of their ability to identify molecular markers or candidate genes associated with economic traits (Yu et al., 2015; Peng et al., 2016). Three reports in this theme attempted to develop genetic linkage maps for three economically important species. Guo et al. used 2b-RAD sequencing technology to construct a high-density genetic

linkage map for a red swamp crayfish (*Procambarus clarkii*) with full-sib family containing 4,878 SNP markers and 94 linkage groups (LGs). Using this genetic map, researchers located the growth- and sex-based QTLs of *P. clarkii* for the first time. Ultimately, 28 QTLs associated with multiple growth traits were identified in nine LGs and a sex-determining major locus was identified in LG20. Notably, the majority of the identified SNPs were female heterozygotes, suggesting that *P. clarkii* may have a ZZ/ZW sex determination system. Similarly, Zhang et al. used a 250 K SNP array to genotype the full-sib F1 family of mirror carp (*Cyprinus carpio*), resulting in the construction of a high-density and high-resolution genetic linkage map for common carp containing 28,416 SNP markers with 50 linkage groups. In this study, a total of 17 QTLs related to feed conversion efficiency (FCE) were mapped on four LGs, and nine candidate genes related to carp FCE involved in multiple biological processes were further identified. In another research, Yu et al. constructed consensus and sex-specific (female and male) genetic linkage maps for large yellow croaker (*Larimichthys crocea*) using samples from an F1 family produced by crossing two distant strains (Daiqu female × Mindong male), which contained 20,147, 11,838, and 11,684 SNPs respectively. Then, the genetic linkage map with 9,885 SNPs and 24 LGs was adapted and integrated based on the physical map of large yellow croaker, and the recombination patterns of each linkage group and the recombination rate of the sex-specific integration map were analyzed in detail. Since previous SNP linkage maps were constructed only for the Mindong strain, this hybrid strain-based linkage map effectively amplified the genetic molecular markers in large yellow croaker.

Theme 4 (Reference genome assembly): Fully assembled and annotated genome will contribute to the study of taxonomic identification, phylogenetic analysis, physiological mechanism and species protection, *etc.* Chen et al. used PacBio and Hi-C sequencing data to construct a first draft of the genome of the pedunculate barnacle (*Capitulum mitella*) with a size of 463. 09 Mb, and generated 16 chromesomes anchored by the contigs. After exhaustive annotation of the genome, the authors also carried out an evolutionary and phylogenetic analysis between *C. mitella* and nine other species based on amino acid sequences.

Theme 5 (Exploratory improvement methods for molecular breeding): Innovative improvement of genotyping methods may bring great improvement to aquaculture molecular breeding. Dagnachew et al. expected to reduce the breeding costs of aquaculture genomic selection by grouping individual samples from reference populations into DNA sample poolings for genotyping, rather than for each individual. Using databases from two salmonid alphavirus-challenged Atlantic salmon (*Salmo salar*) populations, the researchers assessed the impact of the number of DNA poolings in the reference population vs. sequencing coverage on the accuracy of calculating allele frequencies, SNP effects and genomic breeding values (GEBV). A reasonable DNA pooling strategy (i.e., the number of individuals merged in the pool) can provide computational accuracy with little or no loss compared to individual genotyping. And the need for sequencing depth will decrease as the number of DNA pools increases.

Theme 6 (Quantitative genetic basis and genomic architecture of commercial traits): High mortality rates are limiting the Atlantic salmon industry globally. Smoltification is a series of behavioural, developmental and physiological changes during the transfer of Atlantic salmon from freshwater to seawater, which determine the ability of Atlantic salmon to adapt to the seawater environment and therefore urgently needs to be exhaustively investigated. Khaw et al. explored genetic variation of different several phenotypic traits associated with smoltification status by controlling the light regime to simulate seasonal sunshine, and assessed the genetic correlation of smoltification traits between two different age groups from the same family. Conclusion of the study emphasises the importance of the correct choice of time point for measuring phenotypes and fish transfer to seawater, as the smoltification period has different temporal progression between age groups. In the next report, Barría et al. make an outstanding contribution to genetic breeding for feed efficiency in Nile tilapia (*Oreochromis niloticus*). The genetic structure of the population feed efficiency traits was analyzed in detail, suggesting that feed efficiency traits have significant heritability to be used for genetic improvement in Nile tilapia. In addition, genome-wide association analysis (GWAS) was used to locate QTLs for four feed efficiency traits.

Theme 7 (Functional gene impact on omics): Domingues et al. combined computer-assisted sperm analysis and transcriptome analysis to demonstrate that overexpression of growth hormone (*gh*) genes altered the microRNA expression profile of zebrafish testis, thus reducing the sperm motility and reproductive potential. The researchers compared the microRNA transcriptomes of *gh*-transgenic zebrafish (F0104 strain) and non-transgenic zebrafish. 22 differentially expressed microRNAs were monitored, and their functions were adequately interpreted or predicted. Notably, these identified candidate microRNAs may be able to serve as cross-species biomarkers for the validation of male fertility. This study provides the first microRNAome perspective to explain that miRNAs act as direct players in mediating the adverse effects of GH on male reproductive function.

In conclusion, molecular marker-assisted breeding, particularly genomic selection breeding, has great potential to accelerate the rate of genetic gain for excellent characters, which is widely recognized by aquaculture breeders. This Research Topic provides an excellent Research Topic of comprehensive content on aquatic genetics and molecular breeding. We expect

that the continuous improvement of new and high-tech in the future will further accelerate the process of molecular breeding of economical aquatic species.

Author contributions

YY prepared the draft editorial, Z-XG, AH, QL, and LZ improved the manuscript. All authors approved the submission.

Acknowledgments

We would like to thank all the authors and reviewers who have made a valuable contribution to ensuring high quality articles on this Research Topic.

References

Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 557–572. doi:10.1098/rstb.2007.2170

FAO (2020). "The state of world fisheries and aquaculture 2020," in *The state of world Fisheries and aquaculture 2020* (Rome: FAO). doi:10.4060/ca9229en

Gephart, J. A., Henriksson, P. J. G., Parker, R. W. R., Shepon, A., Gorospe, K. D., Bergman, K., et al. (2021). Environmental performance of blue foods. *Nature* 597, 360–365. doi:10.1038/s41586-021-03889-2

Peng, W., Xu, J., Zhang, Y., Feng, J., Dong, C., Jiang, L., et al. (2016). An ultra-high density linkage map and QTL mapping for sex and growth-related traits of common carp (*Cyprinus carpio*). *Sci. Rep.* 6, 26693. doi:10.1038/srep26693

Yu, Y., Zhang, X., Yuan, J., Li, F., Chen, X., Zhao, Y., et al. (2015). Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific white shrimp *Litopenaeus vannamei*. *Sci. Rep.* 5, 15612. doi:10.1038/srep15612

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

frontiers in Genetics

ORIGINAL RESEARCH
published: 31 May 2021
doi: 10.3389/fgene.2021.675928

Transcriptome Profiling Analysis of the Testis After Eyestalk Ablation for Selection of the Candidate Genes Involved in the Male Sexual Development in *Macrobrachium nipponense*

Shubo Jin[1], Yin Fu[2], Yuning Hu[3], Hongtuo Fu[1,3]*, Sufei Jiang[1], Yiwei Xiong[1], Hui Qiao[1], Wenyi Zhang[1], Yongsheng Gong[1] and Yan Wu[1]*

[1] *Key Laboratory of Freshwater Fisheries and Germplasm Resources Utilization, Ministry of Agriculture, Freshwater Fisheries Research Center, Chinese Academy of Fishery Sciences, Wuxi, China,* [2] *Key Laboratory of Marine and Estuarine Fisheries, Ministry of Agriculture, East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai, China,* [3] *Wuxi Fisheries College, Nanjing Agricultural University, Wuxi, China*

OPEN ACCESS

Edited by:
Zexia Gao,
Huazhong Agricultural University, China

Reviewed by:
Shengming Sun,
Shanghai Ocean University, China
Jitao Li,
Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences (CAFS), China

***Correspondence:**
Shubo Jin
jinsb@ffrc.cn
Hongtuo Fu
fuht@ffrc.cn

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

Received: *04 March 2021*
Accepted: *21 April 2021*
Published: *31 May 2021*

Citation:
Jin S, Fu Y, Hu Y, Fu H, Jiang S, Xiong Y, Qiao H, Zhang W, Gong Y and Wu Y (2021) Transcriptome Profiling Analysis of the Testis After Eyestalk Ablation for Selection of the Candidate Genes Involved in the Male Sexual Development in Macrobrachium nipponense. Front. Genet. 12:675928. doi: 10.3389/fgene.2021.675928

The eyestalk of crustacean species secretes many hormones, affecting the process of reproduction, molting, metabolism of glucose, and other functions in crustaceans. In this study, important metabolic pathways and candidate genes involved in the male sexual development were identified through performing the transcriptome profiling analysis of the testis after the ablation of eyestalk from *Macrobrachium nipponense*. The histological observations revealed that the testis development became vigorous after eyestalk ablation, indicating that the hormones secreted by the eyestalk have negative effects on the testis development in *M. nipponense*. Transcriptome profiling analysis revealed that 1,039, 1,226, and 3,682 differentially expressed genes (DEGs) were identified between normal prawns (CG) vs single-side eyestalk ablation prawns (SS), SS vs double-side eyestalk ablation prawns (DS), and CG vs DS, respectively, indicating that the ablation of double-side eyestalk has more significant regulatory roles on male sexual development than that of single-side ablation, which was consistent with the histological observations. Lysosome, Apoptosis, Glycolysis/Gluconeogenesis, and Insulin signaling pathway were the main enriched metabolic pathways in all of these three comparisons, and the important genes from these metabolic pathways were also selected. The qPCR verifications of 10 DEGs from these metabolic pathways were the same as those of RNA-seq. The qPCR, *in situ* hybridization, and RNA interference analysis of Mn-$NF_kB\alpha$ revealed that $NF_kB\alpha$ has a positive regulatory effect on testis development. This study provided new insights on male sexual development in *M. nipponense*, promoting the studies on male sexual development in other crustaceans as well.

Keywords: *Macrobrachium nipponense*, eyestalk ablation, testis, male sexual development, $NF_kB\alpha$

INTRODUCTION

The oriental river prawn, *Macrobrachium nipponense* (Crustacea; Decapoda; and Palaemonidae), is widely distributed in China and other Asian countries (Cai and Shokita, 2006; Salman et al., 2006; Ma et al., 2011), which is an important commercial species with annual aquaculture production that reached 205,010 tons in 2016 (Wang et al., 2019). The same as other *Macrobrachium* species, male prawns grow faster and reach larger size at the harvest time (Ma et al., 2011). Thus, male prawns are preferred in the *M. nipponense* aquaculture. But the rapid development of the testis in the reproductive season is another main problem that restricted the sustainable development of *M. nipponense*. Previous studies revealed that the testis of a newborn *M. nipponense* can reach sexual maturity within 40 days after hatching (Jin et al., 2016). Thus, inbreeding will happen between the newborn prawns. Inbreeding will lead to the decrease of the ability of resistance to adversity in their offspring, the small scale of market prawn, and the degradation of germplasm resources. Therefore, it is urgently needed to fully understand the male sexual determination and development mechanism, especially for the identification of the key metabolic pathways and genes involved in the mechanism of male sexual determination and development in *M. nipponense*, with the aims of establishing the technique to produce all male progeny on a commercial scale and regulating the process of testis development.

The eyestalk of crustacean species has many neurosecretory structures. The X-organ–SG complex (XO–SG) is located in the eyestalk in crustaceans, which was identified as a principal neuroendocrine gland (Hopkins, 2012). It stores and releases the crustacean hyperglycemic hormone (CHH) superfamily neurohormones, including CHH, gonad-inhibiting hormone (GIH), molt-inhibiting hormone (MIH), ion transport peptides, and mandibular organ-inhibiting hormone (MOIH), playing essential roles in reproduction (Treerattrakool et al., 2011, 2013; Revathi et al., 2013), molting (Pamuru et al., 2012; Salma et al., 2012; Shen et al., 2013), metabolism of glucose (Santos et al., 1997; Almeida et al., 2004), and other functions (Tiu and Chan, 2007; Sainz-Hernández et al., 2008; Diarte-Plata et al., 2012). Knockdown of the expression of GIH by RNA interference (RNAi) promotes the ovarian development in *M. nipponense* (Qiao et al., 2015). Knockdown of the expression of MIH by RNAi promotes the molting in *M. nipponense* (Qiao et al., 2018). CHH has been proven to promote testis development in *M. nipponense* (Jin et al., 2013b). Many previous studies have proven that the ablation of eyestalk from crustacean species has positive effects on the expression of insulin-like androgenic hormone (IAG; Sroyraya et al., 2010; Chung et al., 2011; Guo et al., 2019). IAG is an important gene playing essential roles in the male sexual differentiation and development in many crustacean species (Ventura et al., 2009, 2011). Knockdown of the expression of IAG resulted in sex reversal in *Macrobrachium rosenbergii*, and the all-male progeny were produced when the "reversal females" were mating with normal male *M. rosenbergii* (Ventura et al., 2012).

It is widely acknowledged that the testis plays essential regulatory roles in reproduction, sexual maturity, and sex differentiation. Previous male reproductive studies have been conducted at the molecular and cellular levels in *M. nipponense* (Qiu et al., 1995; Yang et al., 1999; Cao, 2006; Guo, 2007). A total of 52 candidate male reproduction-related genes were identified from the testis cDNA library of *M. nipponense* (Qiao et al., 2012). An integrated analysis of metabolomes and transcriptomes was also performed in the testis between the reproductive season and non-reproductive season, in order to select candidate male reproduction-related metabolites and genes, regulated by different water temperature and illumination time (Jin et al., 2020). Some candidate genes from these constructed testis transcriptomes have proven their functions in the mechanism of male sexual differentiation and development (Zhang et al., 2013a,b,c). These studies dramatically improved the studies on the mechanism of male sexual differentiation and development in *M. nipponense*. However, the effects of eyestalk on male sexual differentiation and development were still unclear.

In this study, we aimed to select the vital metabolic pathways and genes involved in the male sexual differentiation and development in *M. nipponense* through performing the transcriptome profiling analysis of the testis after the ablation of single-side and double-side eyestalks. The functions of nuclear factor kappa B inhibitor alpha ($NF_kB\alpha$) were further analyzed in-depth by using qPCR analysis, *in situ* hybridization, and RNAi. This study provided valuable evidences on the studies of male sexual differentiation and development in *M. nipponense*, as well as other crustacean species.

MATERIALS AND METHODS

Ethics Statement

We obtained the permission from the committee of Freshwater Fisheries Research Center and the Tai Lake Fishery Management Council during the experimental programs. MS222 anesthesia was used to sedate the prawns and shear the tissues.

Sample Collection

A total of 600 healthy male *M. nipponense* prawns were collected from a wild population in Tai Lake, Wuxi, China (120°13′44″E, 31°28′ 22′N) with body weights of 3.34–4.76 g. All the samples were randomly divided and transferred to three 500-L tanks and maintained in aerated freshwater for 3 days with dissolved oxygen of ≥6 mg/L prior to the tissue collection. The three groups were normal prawns (CG), single-side eyestalk ablation prawns (SS), and double-side eyestalk ablation prawns (DS). The death rate of DS prawns was 13.7 and 8.6% higher than that of CG prawns and SS prawns, respectively. The testis was collected from these three groups after 7 days of treatment and immediately preserved in liquid nitrogen until use for transcriptomic analysis.

Transcriptomic Profiling Analysis

The comparative transcriptome analysis of androgenic gland between the CG, SS, and DS was performed. In order to ensure the sufficient amount of RNA samples, the testis from at least 30

prawns were pooled to form one biological replicate, and three biological replicates were sequenced for all of these three groups. Thus, a total of nine libraries were generated for sequencing. The experimental process of transcriptome sequencing has been well described in the previously published studies (Jin et al., 2013a, 2017, 2020). Briefly, the total RNA from each pooled sample was extracted by using RNAiso Plus Reagent (TaKaRa), following the manufacturer's instructions. The concentration of total RNA was measure by a spectrophotometer (Eppendorf), and the integrity was measured by using a 2100 Bioanalyzer (Agilent Technologies, Inc.) with a minimum RNA integrity number (RIN) value of 7.0. A total of 4 μg of total RNA was used to construct the library, and Illumina HiSeq 2500 sequencing platform was used to perform the sequencing under the parameter of PE150.

Raw data of fastq format were firstly processed using Trimmomatic with default parameters (Bolger et al., 2014). The clean reads were assembled into expressed sequence tag clusters (contigs) and *de novo* assembled into transcripts by Trinity (version 2.4) with paired-end method with default parameters after removing the adaptor and low-quality sequences (Grabherr et al., 2011). The gene annotation was then performed in the NR protein, prior to Gene Ontology (GO), the Clusters of Orthologous Groups of proteins (COG), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses, using an *E*-value cutoff of 10^{-5} (Jin et al., 2013a). GO (Ashburner et al., 2000), COG (Tatusov et al., 2003), and KEGG (Minoru et al., 2008) analyses were annotated by using Blast2go software and Blast software. The criteria of false discovery rate < 0.05 was used to filter the differentially expressed genes (DEGs) by EB-seq algorithm (Benjamini et al., 2001).

qPCR Analysis

A total of 15 male and female prawns were collected from a wild population in Tai Lake with body weights of 3.12–3.87 and 1.97–2.54 g, respectively, in order to obtain tissues for qPCR analysis in different mature tissues. All the samples were transferred to 500-L tanks and maintained in aerated freshwater for 3 days with dissolved oxygen of ≥6 mg/L prior to the tissue collection. Different mature tissues included the testis, ovary, hepatopancreas, muscle, eyestalk, gill, heart, and brain ($N = 5$). Specimens for the different stages of post-larval developmental stages were from the full-sibs population, collected during their maturation process ($N = 5$).

qPCR was performed on the Bio-Rad iCycler iQ5 Real-Time PCR System (Bio-Rad), which was used to carry out the SYBR Green RT-qPCR assay. The procedure has been well described in detail in previous studies (Zhang et al., 2013a,b,c). Briefly, the total RNA from each tissue was extracted by using RNAiso Plus Reagent (TaKaRa), following the manufacturer's instructions. The concentration of total RNA was measured by a spectrophotometer (Eppendorf), and the integrity was measured by agarose gel. Approximately 1 μg of total RNA from each tissue was used for first-strand cDNA synthesis by using iScript^{TM} cDNA Synthesis Kit Perfect Real Time (BIO-RAD), following the manufacturer's instructions. Amplifications were performed in a 96-well plate with a 25-μl reaction volume containing 12.5 μl of 2 × Ultra SYBR Mix (CWBIO), 0.5 μl of each primer, 1 μl of cDNA template, and 10.5 μl of PCR-grade water. The thermal profile for qPCR was 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Each tissue was performed in triplicate. The relative gene expression was calculated based on the $2^{-\Delta\ \Delta\ CT}$ comparative CT method (Livak and Schmittgen, 2001). The primers used for qPCR verification of important DEGs are listed in **Table 1**. The primers used for qPCR analysis of Mn-$NF_kB\alpha$ are listed in **Table 2**. EIF was used as the reference gene in this study (Hu et al., 2018). Different concentrations of testis cDNA templates were used to measure the amplification efficiency of Mn-$NF_kB\alpha$ and EIF, including undiluted, two times diluted, four times diluted, and eight times diluted samples. The slope of the Mn-$NF_kB\alpha$ and EIF at different concentrations of diluted samples was 1.412 and 1.423, respectively, indicating that the amplification efficiency between the Mn-$NF_kB\alpha$ and EIF is the same in this study.

In situ Hybridization

The mRNA locations of Mn-$NF_kB\alpha$ in the testis, androgenic gland, and different reproductive cycles of ovary were analyzed by using *in situ* hybridization. The different reproductive cycles of ovary were collected, according to the previous study (Qiao et al., 2015). The testis and androgenic gland were collected in reproductive season. Primer5 software was used to design the anti-sense and sense probes of chromogenic *in situ* hybridization study and synthesized with DIG signal by Shanghai Sangon Biotech Company. The sequences of anti-sense and sense probes are listed in **Table 1**. The previous

TABLE 1 | Primers used for qPCR verification.

Primer	Sequence
Sialin-F	ATCAAAGGAATGTCTGCTACCGT
Sialin-R	TCAGGTAAATCGTTCCAGGGATG
Alpha-F	CAACGACTTTGTCACCAGGAAAA
Alpha-R	TGGTATTCCCTGACCCCATCTAT
ASK1-F	GAATTCTCTCGGAGCATATCCGT
ASK1-R	TCTTCAGGAGGTAGAACCCATCT
$NF_kB\alpha$-F	CATGGTGACCCAGTTAACCAGA
$NF_kB\alpha$-R	CGTCAAGTGTTGCAGGATTTCTT
TGF-F	CATCTTTACCAGAGTGTGTGGGA
TGF-R	CTGCTTACGAATACCCTGTTCCT
ADP-F	CACACCGATGTTTACTTCTGGGA
ADP-R	CAACACATGATCTCCTGGCTGAA
PPT-F	GCACTTGGAAGACGAGATGATTG
PPT-R	GGAACCGTGAGTTTGTAGCTTTC
Hexokinase-F	CACAGGATGCTTCTTTGGAGGA
Hexokinase-R	GAGAGTCTTCCCCTGAATCAAGA
Alcohol-F	TAAAACACCATCCCCCAGAGAAG
Alcohol-R	AGGGTAAGTTTGGATCCCTCAAC
Acetyl-F	CCAAGCTTCAGAACGGATACAAC
Acetyl-R	GACGAAACCACCATTCAAAGAGG

TABLE 2 | Primers used for NFkB analysis.

Primer name	Nucleotide sequence (5′→3′)	Purpose
$NF_kB\alpha$-RTF	CATGGTGACCCAGTTAACCAGA	FWD primer for $NF_kB\alpha$ expression
$NF_kB\alpha$-RTR	CGTCAAGTGTTGCAGGATTTCTT	RVS primer for $NF_kB\alpha$ expression
EIF-F	CATGGATGTACCTGTGGTGAAAC	FWD primer for EIF expression
EIF-R	CTGTCAGCAGAAGGTCCTCATTA	RVS primer for EIF expression
$NF_kB\alpha$ anti-sense Probe	TTCGCGAAAGAGGGAGAGGAGGAATCAGCGATCCCT	Probe for $NF_kB\alpha$ ISH analysis
$NF_kB\alpha$ sense Probe	AGGGATCGCTGATTCCTCCTCTCCCTCTTTCGCGAA	Probe for $NF_kB\alpha$ ISH analysis
$NF_kB\alpha$ RNAi-F	TAATACGACTCACTATAGGGGAAAGAGGAAGGCTCCGTTA	FWD primer for RNAi analysis
$NF_kB\alpha$ RNAi-R	TAATACGACTCACTATAGGGGGTCCTTCCGGTATGATTTG	RVS primer for RNAi analysis

studies have described well the detailed procedures (Jin et al., 2018; Li et al., 2018). Slides were examined under a light microscope for evaluation.

RNA Interference Analysis

RNA interference was performed to analyze the regulatory roles on Mn-$NF_kB\alpha$ in *M. nipponense*. The specific RNAi primer with T7 promoter site was designed by using Snap Dragon tools[1] and is shown in **Table 1**. The Transcript Aid™ T7 High Yield Transcription kit (Fermentas Inc., United States) was used to synthesize the Mn-$NF_kB\alpha$ dsRNA, followed by the procedures of the manufacturer. A total of 300 healthy mature male *M. nipponense* were collected with body weight of 3.17–4.96 g and divided into two groups. As described in previous studies (Jiang et al., 2014; Jin et al., 2018), the prawns from experimental group were injected with 4 μg/g of Mn-$NF_kB\alpha$ dsRNA, while the prawns from the control group were injected with an equal volume of green fluorescent protein. The $NF_kB\alpha$ mRNA expression was investigated in the androgenic gland by qPCR after the injection at 1, 7, and 14 days in order to detect the interference efficiency ($N \geq 5$). The mRNA expressions of Mn-IAG were also measured in the androgenic gland templates from the same prawns in order to analyze the regulatory relationship between Mn-$NF_kB\alpha$ and Mn-IAG.

Histological Observation

The morphological changes of the testis between different days after RNAi treatment were observed by hematoxylin and eosin (H&E) staining. Five testicular samples were collected after 1, 7, and 14 days of RNAi treatment for H&E staining. The procedures have been described well in previous studies (ShangGuan et al., 1991; Ma et al., 2006). Olympus SZX16 microscope was used to observe the slides (Olympus Corporation, Tokyo, Japan). The various cell types were labeled based on morphological analysis (Jin et al., 2016).

Statistical Analysis

SPSS Statistics 23.0 was used to measure the statistical differences, estimated by one-way ANOVA followed by least significant difference and Duncan's multiple range test. Quantitative data were expressed as mean ± SD. $p < 0.05$ indicates a significant difference.

[1]http://www.flyrnai.org/cgibin/RNAifind_primers.pl

RESULTS

Histological Observations of the Testis After Eyestalk Ablation

Histological observations were performed in order to analyze the morphological changes of the testis after the ablation of eyestalk in *M. nipponense* (**Figure 1**). The histological observations revealed that the greatest number of spermatogonia was observed in the CG prawns, followed by SS prawns and DS prawns. However, the dominant cells in the DS prawns were sperms, which were more than those in SS prawns and CG prawns. Spermatogonia were rarely observed in the DS prawns.

Transcriptome Analysis

The transcriptome generated 54,341 non-redundant transcripts with an average length of 1,311.61 bp. The non-redundant transcripts length ranged from 301 to 28,887 bp. The majority of the transcripts was 301–400 bp (23.62%) in length, followed by >2,000 bp (19.61%) and 401–500 bp (13.36%). The complete and duplicated BUSCOs of this assembled transcriptome reached 97.5%, indicating the completeness of this assembled transcriptome.

All of the assembled unigenes were firstly annotated in the Nr (non-redundant) database. A total of 17,660 (32.50%)

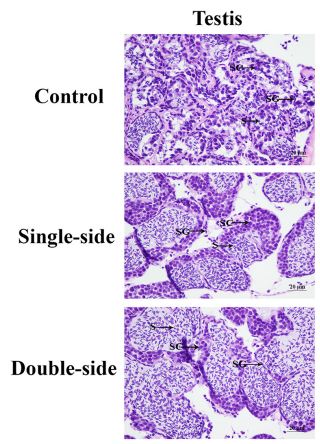

FIGURE 1 | The morphological differences of the testis after the ablation of eyestalk. SG, spermatogonia; SC, spermatocytes; S, sperms; and CT, collected tissue. Scale bars = 20 μm.

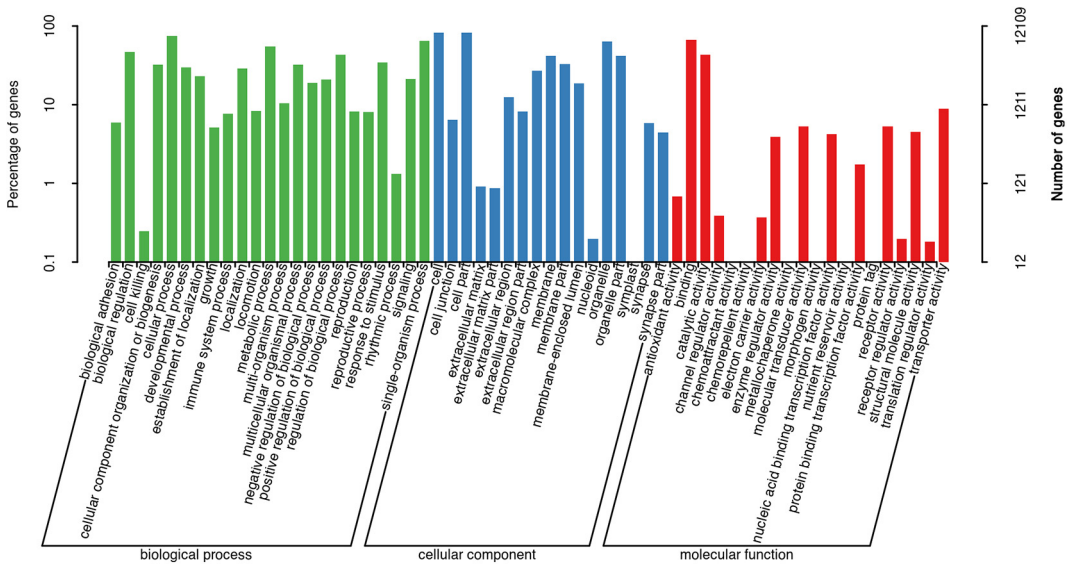

FIGURE 2 | Gene ontology classification of non-redundant transcripts.

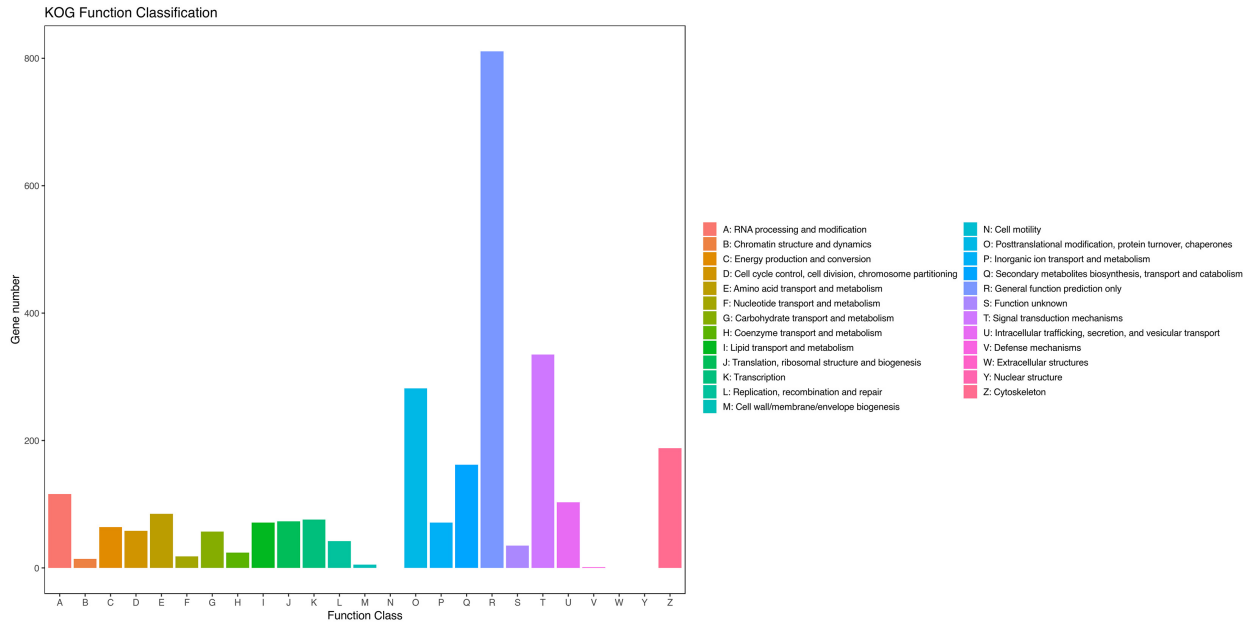

FIGURE 3 | Clusters of orthologous groups of proteins (COG) classification of putative proteins.

unigenes were annotated in the Nr database, while the other unannotated unigenes represent novel genes, but the functions need further investigations.

The assembled unigenes were then annotated in the GO, COG, and KEGG databases. GO and COG analyses provide a structured vocabulary to describe the transcripts. A total of 12,109 unigenes matched the known proteins in GO database, composed of 60 functional groups (**Figure 2**). The number of unigenes in each functional group ranged from 1 to 10,057. Cell, Cell part, Cellular process, and Binding represent the main functional groups, in which the number of unigenes was >8,000. A total of 2,509 unigenes were assigned to the matched proteins in COG database, including 22 functional categories (**Figure 3**). The unigenes in each functional category ranged from 1 to 811. The main functional category includes General function prediction only, Signal transduction mechanisms, and Posttranslational modification, protein turnover, and chaperones, in which the number of unigenes was more than 200.

Kyoto Encyclopedia of Genes and Genomes analysis plays essential roles in releasing the regulatory relationship between

the unigenes, assembled in this transcriptome. A total of 4,971 unigenes were matched with the known proteins in the KEGG database involved in the 338 metabolic pathways. The metabolic pathways, in which the number of unigenes was more than 200, included Alzheimer disease, Pathways in cancer, and Huntington disease.

Identification of Differentially Expressed Genes

The DEGs were identified using the criterion of >2.0 as up-regulatory genes and <0.5 as down-regulatory genes, and p-value < 0.05. A total of 1,039 DEGs were identified between CG and SS, including 617 up-regulated genes and 422 down-regulated genes. Eighty-seven metabolic pathways were identified, and the number of DEGs in each metabolic pathway ranged from 1 to 4. A total of 1,226 DEGs were identified between SS and DS, including 739 up-regulated genes and 487 down-regulated genes. A total of 196 metabolic pathways were identified, and the number of DEGs in each metabolic pathway ranged from 1 to 8. A total of 3,682 DEGs were found between CG and DS, including 1,978 up-regulatory genes and 1,704 down-regulatory genes. A total of 285 metabolic pathways were identified, and the number of DEGs in each metabolic pathway ranged from 1 to 56. KEGG analysis revealed that Lysosome, Apoptosis, Insulin signaling pathway, and Glycolysis/Gluconeogenesis were the main enriched metabolic pathways in all of these three comparisons.

Ten important DEGs were identified from these metabolic pathways, which were differentially expressed in at least two comparisons (**Table 3**). Sialin-like, alpha-L-fucosidase, and acetyl-CoA carboxylase (ACC) were selected from the metabolic pathway of Lysosome. Apoptosis signal-regulating kinase 1 (ASK1), $NF_kB\alpha$, and TGF-beta-activated kinase 1 (TGF) were selected from the metabolic pathway of Apoptosis. Alcohol dehydrogenase class-P (ADP), Palmitoyl-protein thioesterase 1 (PPT1), and Hexokinase (HXK) were selected from the metabolic pathway of Glycolysis/Gluconeogenesis.

qPCR Verification of Important Differentially Expressed Genes

The expressions of 10 important DEGs were verified by qPCR, which showed the same expression pattern with that of RNA-Seq (**Figure 4**). The expressions of $NF_KB\alpha$ and PPT1 were gradually increased from the control group to double-side ablation and showed a significant difference between each group ($p < 0.05$). The lowest expression of sialin-like was observed in the control group and showed a significant difference with that of single-side ablation and double-side ablation ($p < 0.05$), while the highest expressions of alpha-L-fucosidase, TGF, HXK, and ACC were observed in the control group and showed a significant difference with that of single-side ablation and double-side ablation ($p < 0.05$). The highest expression of Alcohol dehydrogenase class II was observed in the double-side ablation and showed a significant difference with that of the control group and single-side ablation ($p < 0.05$), which the lowest expressions of ASK1 and ADP were observed in double-side ablation and showed a significant difference with that of the control group and single-side ablation ($p < 0.05$).

qPCR Analysis of Mn-$NF_kB\alpha$

The physiology functions of a gene can be reflected by the qPCR analysis. According to the qPCR analysis in different tissues,

TABLE 3 | Important DEGs through transcriptome profiling analysis.

Name	Accession number	*p* value	CG vs SS	CG vs DS	SS vs DS	Metabolic pathways
				Fold change		
Sialin-like	XP_018006493.1	0.0367	0.32	0.39		Lysosome
Alpha-L-fucosidase	KFM70007.1	0.0105	2.01	2.23		Lysosome; other glycan degradation
Apoptosis signal-regulating kinase 1 (ASK1)	AKI88007.1	0.0145		2.21	2.13	Apoptosis; platinum drug resistance; tight junction
NF-kappa B inhibitor alpha ($NF_kB\alpha$)	AET34918.1	0.0161	0.46	0.28	0.43	Apoptosis; Pathways in cancer; shigellosis
TGF-beta-activated kinase 1 (TGF)	AKV88638.1	0.0338	2.31	2.17		Apoptosis; NF-kappa B signaling pathway; Shigellosis; MAPK signaling pathway
Alcohol dehydrogenase class-P (ADP)	XP_019577393.1	0.0293		2.43	2.16	Glycolysis/Gluconeogenesis; fatty acid degradation; tyrosine metabolism
Palmitoyl-protein thioesterase 1 (PPT1)	XP_015686279.1	0.0082		0.27	0.31	Glycolysis/Gluconeogenesis; shigellosis
Hexokinase	ABO21409.1	2.58E-06	3.23	3.32		Glycolysis/Gluconeogenesis; fatty acid degradation; drug metabolism—cytochrome P450
Alcohol dehydrogenase class II (Alcohol)	CCQ25768.1	0.0275		0.45	0.48	Insulin signaling pathway; pyruvate metabolism; AMPK signaling pathway
Acetyl-CoA carboxylase (Acetyl)	ALK82309.1	8.70E-05	3.23	3.13		Lysosome; fatty acid elongation

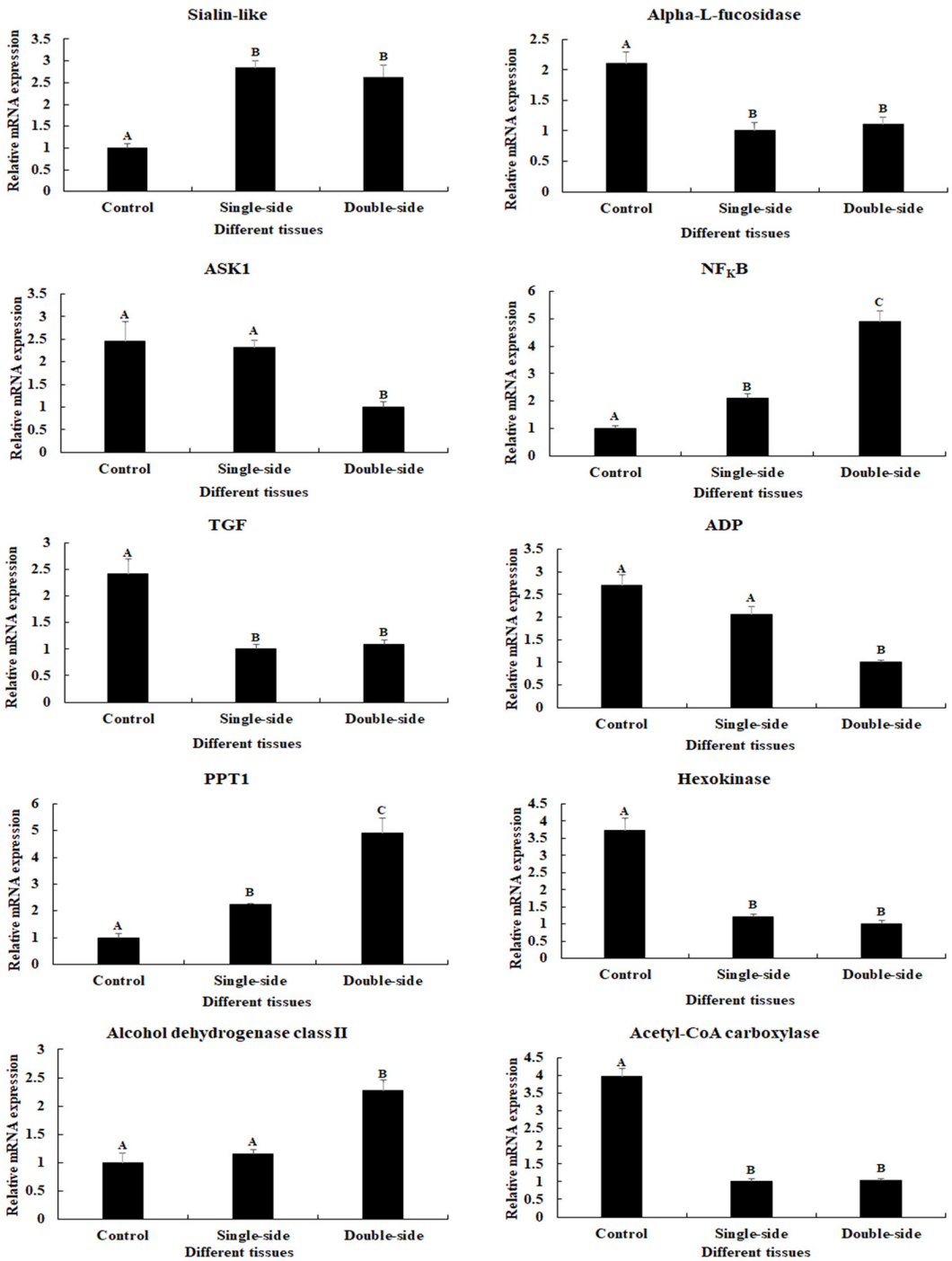

FIGURE 4 | Verification of the expressions of 10 differentially expressed genes (DEGs) by qPCR. The amounts of DEG expression were normalized to the EIF transcript level. Data are shown as mean ± SD (standard deviation) of tissues in three separate individuals. Capital letters indicate expression difference. Control indicates normal prawns; single-side indicates the single-side ablation of eyestalk; double-side indicates the double-side ablation of eyestalk.

the highest expression of Mn-NF$_k$Bα was observed in the testis, which was significantly higher than the other tested tissues and showed a significant difference ($p < 0.05$; **Figure 5A**). The lowest expression was observed in the brain. The expression in the testis was 6.58-fold higher than that of the brain.

The mRNA expression of Mn-NF$_k$Bα was also measured during the post-larval developmental stages. The PCR analysis revealed that the expression of Mn-NF$_k$Bα was gradually increased in time with specimen development (**Figure 5B**). The gonad can be distinguished for the first time by the naked eye at

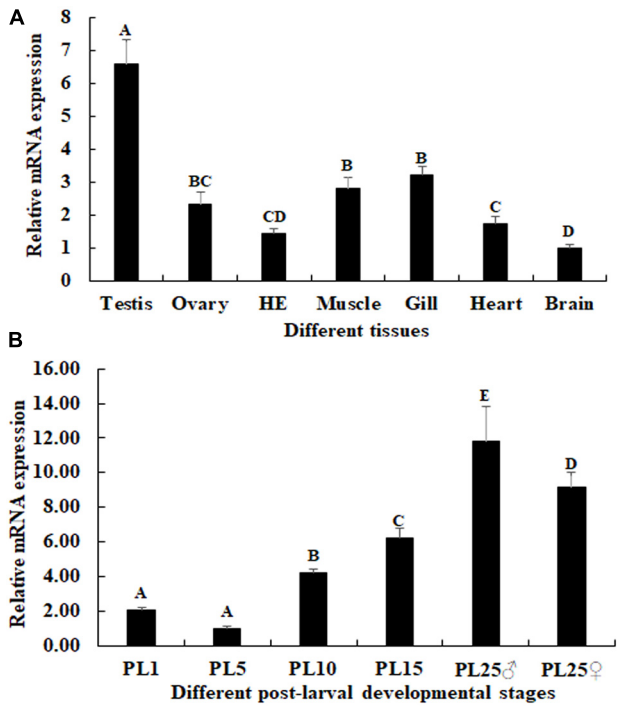

FIGURE 5 | Expression characterization of Mn-NF$_k$Bα in different tissues and post-larval developmental stages. The amount of Mn-NF$_k$Bα mRNA was normalized to the EIF transcript level. Data are shown as mean ± SD (standard deviation) of tissues from three separate individuals. Capital letters indicate expression difference between different samples. **(A)** The expression characterization of Mn-NF$_k$Bα in different tissues. **(B)** The expression characterization of Mn-NF$_k$Bα in different post-larval developmental stages.

PL25. The expression of Mn-NF$_k$Bα was higher at both PL25♂ and PL25♀ and showed a significant difference with that of other developmental stages ($p < 0.05$). However, the expression at PL25♂ was higher than that of PL25♀ ($p < 0.05$). The lowest expression was observed in PL5, and the expressions in PL25♂ and PL25♀ were 11.83- and 9.15-fold higher than those of PL5, respectively.

In situ Hybridization of Mn-NF$_k$Bα

The cell type was labeled, based on the previous study (**Figure 6**). According to the *in situ* hybridization analysis, signals of Mn-NF$_k$Bα were observed in spermatogonia and spermatocytes, whereas no signal was observed in sperms. Strong mRNA signals in the androgenic gland were only observed in the ejaculatory bulb surrounding the androgenic gland cells, while no signals were directly found in all stages of androgenic gland cells (**Figure 6**). Clear signals were rarely observed in O I and O V, while signals were observed in the nucleus, yolk granule, yolk granule, and cytoplasmic membrane in O II, O III, and O IV.

The RNA Interference Analysis of Mn-NF$_k$Bα

The potential functions of Mn-NF$_k$Bα on male sexual development in *M. nipponense* were analyzed by using RNAi. The expression levels of Mn-NF$_k$Bα were measured in the testis after the treatment of Mn-NF$_k$Bα dsRNA. According to the qPCR analysis, the expression of Mn-NF$_k$Bα remained stable in the control group after the injection of GFP and showed no significant difference ($p > 0.05$). However, the expression of Mn-NF$_k$Bα significantly decreased at days 7 and 14 after the injection of Mn-NF$_k$Bα dsRNA. The decrease reached 95 and 85% at days 7 and 14, respectively, compared with that in the control group (**Figure 7A**).

The expressions of Mn-IAG were also measured in the androgenic gland from the same prawns (**Figure 7B**). According to the qPCR analysis, the expression of Mn-IAG at day 1 in the control group was slightly higher than that of day 7 and day 14, while it generally remained stable. In the RNAi group, the expressions of Mn-IAG were significantly decreased at day 7 and day 14 after the injection of Mn-NF$_k$Bα dsRNA. The expression decreased about 61 and 54% at days 7 and 14, respectively, compared with that in the control group.

Histological Observations of the Testis After RNA Interference

According to the histological observations, the number of sperms was more than that of spermatogonia and spermatocytes in the control groups. Compared with that of the control group at day 7 and day 14, the number of sperms in the RNAi group was significantly decreased. In the RNAi group, the number of sperms was gradually decreased in time with Mn-NF$_k$Bα dsRNA treatment, and sperms were rarely found at day 14 after Mn-NF$_k$Bα dsRNA treatment (**Figure 8**).

DISCUSSION

The eyestalk of crustaceans secreted many neurosecretory structures and mediated the reproduction, molting, and metabolism of glucose in crustaceans (Jin et al., 2013b; Qiao et al., 2015, 2018). In this study, we aimed to analyze the regulatory effects on male sexual development through performing the transcriptome profiling analysis of the testis after eyestalk ablation. The histological observations of the testis after eyestalk ablation from *M. nipponense* indicated that the number of sperms in the DS prawns was significantly more than that of SS prawns and CG prawns, and spermatogonia were rarely observed in the DS prawns. This indicated that the hormones secreted by the eyestalk have negative regulatory effects on the testis development. This is the same as the results of a previous study that the hormones secreted by eyestalk inhibit the expression of IAG in *M. nipponense* (Li et al., 2015), and IAG promoted the male sexual characteristic development in many crustacean species (Ventura et al., 2009, 2011).

A total of 54,341 transcripts were generated in this study, providing valuable evidences on the studies of male sexual development. According to the GO and COG analyses, the genes related to the male sexual development were predicted to be mainly found in the functional groups of Cell, Cell part, Cellular process, and Binding in the GO assignment, and in the functional groups of General function prediction

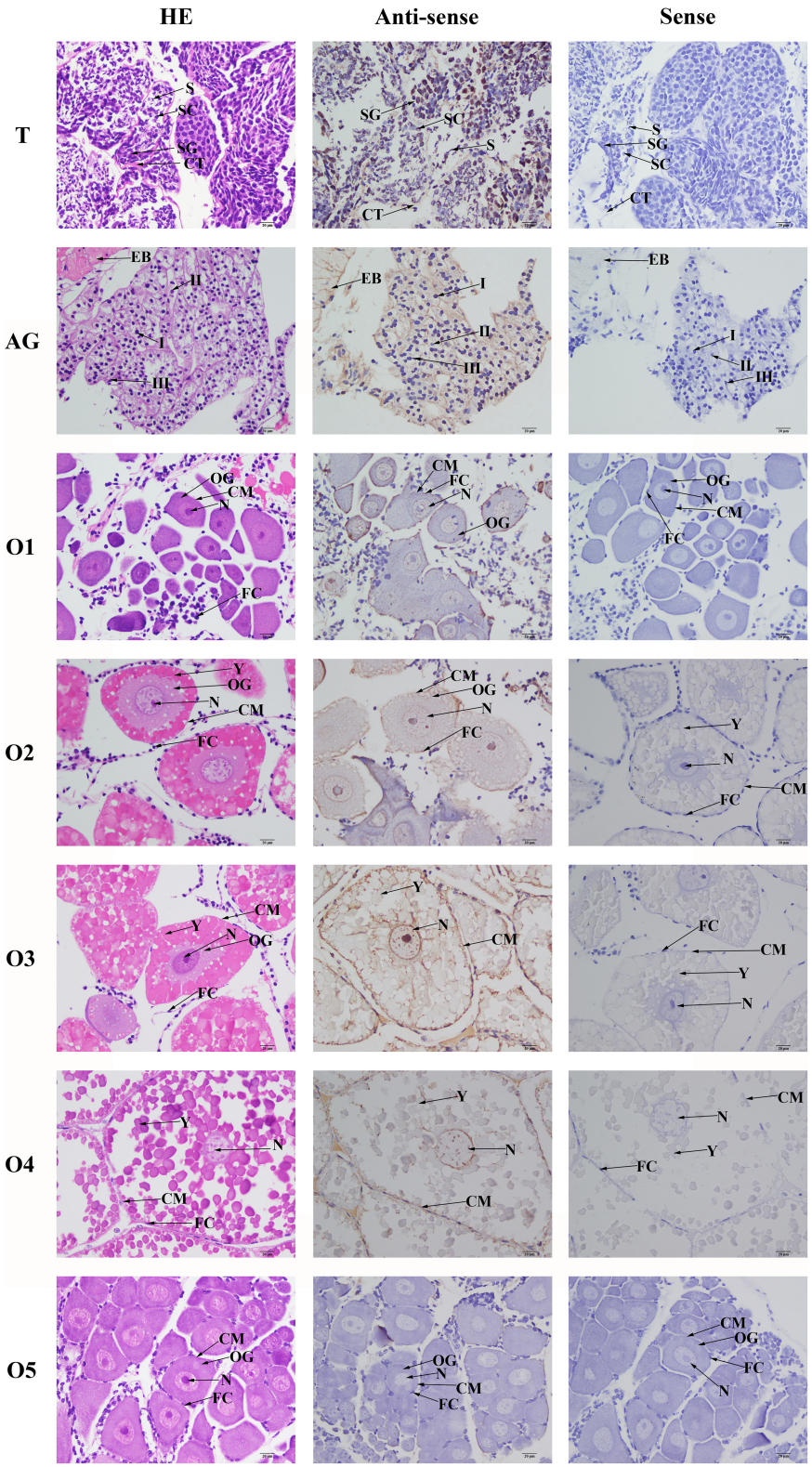

FIGURE 6 | *In situ* hybridization analysis of Mn-NF$_k$Bα gene in the testis and androgenic gland from reproductive season, and different reproductive cycle of ovary of *M. nipponense*. SG, spermatogonia; SC, spermatocytes; S, sperms; CT, collected tissue; I, Stage I of androgenic gland cell; II, Stage II of androgenic gland cell; III, Stage III of androgenic gland cell; EB, ejaculatory bulb; OG, oogonium; OC, oocyte; CM, cytoplasmic membrane; N, nucleus; Y, yolk granule; and FC, follicle membrane. Scale bars = 20 μm.

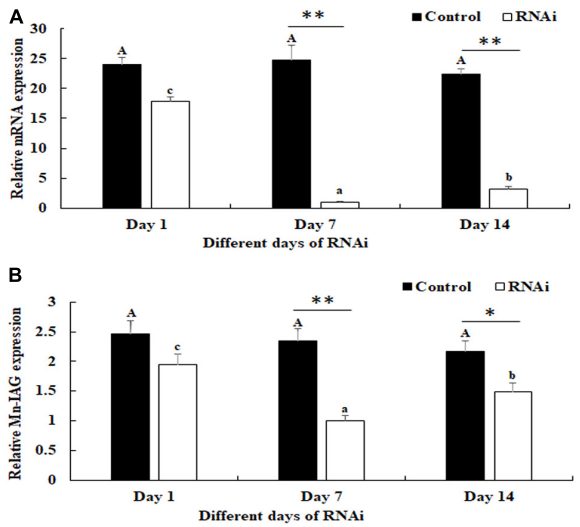

FIGURE 7 | Expression characterization of Mn-$NF_kB\alpha$ and Mn-IAG at different days after Mn-$NF_kB\alpha$ dsRNA injection. The amount of Mn-$NF_kB\alpha$ and Mn-IAG mRNA was normalized to the EIF transcript level. Data are shown as mean ± SD (standard deviation) of tissues from three separate individuals. Capital letters indicate expression difference between different days after green fluorescent protein (GFP) injection in the control group. Lowercase letters indicate expression difference between different days after Mn-$NF_kB\alpha$ dsRNA injection in the RNA interference (RNAi) group. * ($p < 0.05$) and ** ($p < 0.01$) indicate significant expression difference between the RNAi group and control group at the sample day. **(A)** Expression characterization of Mn-$NF_kB\alpha$ at different days after Mn-$NF_kB\alpha$ dsRNA injection. **(B)** Expression characterization of Mn-IAG at different days after Mn-$NF_kB\alpha$ dsRNA injection.

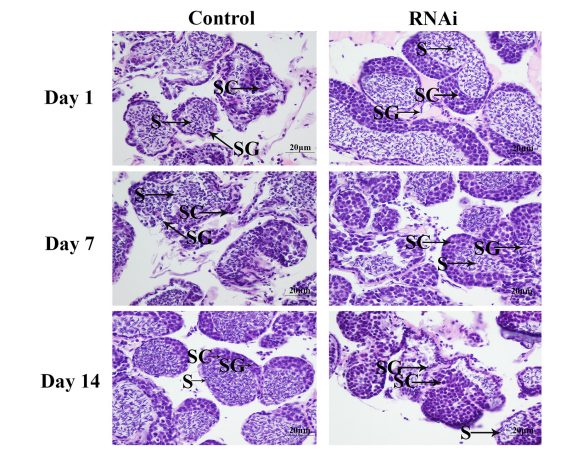

FIGURE 8 | The morphological differences of the testis between the RNA interference (RNAi) and control groups. SG, spermatogonia; SC, spermatocytes; S, sperms; and CT, collected tissue. Scale bars = 20 μm.

only, Signal transduction mechanisms, and Posttranslational modification, protein turnover, and chaperones in the COG classification, which were consistent with the previous studies (Jin et al., 2017, 2020). The number of DEGs between CG vs SS, SS vs DS, and CG vs DS was 1,039, 1,226, and 3,682, respectively, indicating that the ablation of double-side eyestalk has more regulatory effects on male sexual development than the single-side ablation in *M. nipponense*, which was consistent with histological observations of the testis after eyestalk ablation. KEGG analysis revealed that Lysosome, Apoptosis, Insulin signaling pathway, and Glycolysis/Gluconeogenesis were the main enriched metabolic pathways in all of these three comparisons, predicting that these mainly enriched metabolic pathways and vital DEGs from these metabolic pathways may play essential roles in male sexual development in *M. nipponense*. qPCR verification of these DEGs showed the same expression pattern with that of RNA-Seq, indicating the accuracy of the RNA-Seq.

Lysosomes are organelles playing essential roles in decomposing proteins, nucleic acids, polysaccharides, and other biological macromolecules. Lysosomes contain many hydrolases, functioning in the decomposition of the substances that enter into cells from the outside or the digestion of the local cytoplasm or organelles in the cells. The lysosomes will rupture, and the hydrolases will be released to digest the whole cells when cells are aged (Duve and Wattiaux, 1966; Luzio et al., 2007). Apoptosis refers to the programmed cell death, carefully controlled by genes to maintain the stability of internal environment. Apoptosis is an active process, which is different from cell necrosis. The process of apoptosis involves the activation, expression, and regulation of a series of genes, in order to better adapt to the living environment. Apoptosis plays important roles in the mechanism of physiology and pathology, in order to respond various stimuli, including ischemia, hypoxia, exposure to certain drugs and chemicals, immune reactions, infectious agents, high temperature, radiation, and various disease states (Thompson, 1995; Johnstone et al., 2002). According to the histological observations, the testis development was vigorous after eyestalk ablation. Thus, Lysosomes and Apoptosis were needed to digest the aged cells and to adapt to stress, in order to maintain the normal testis development. Alpha-L-fucosidase was a lysosomal enzyme, presented in all mammalian cells. Its activity has been proven to be deficient in the human autosomal recessive disease fucosidosis. It is also proposed as a marker of hepatocellular carcinoma (Fukushima et al., 1985). ACC is the enzyme playing essential roles in the synthesis of malonyl-coenzyme A (malonyl-CoA; Winder and Hardie, 1996). Malonyl-CoA is a key metabolite in the regulation of energy homeostasis. Malonyl-CoA was an inhibitor of fatty acid oxidation in skeletal muscle mitochondria, decreased in rat skeletal muscle during exercise or in response to electrical stimulation (Abu-Elheiga et al., 2001). ASK1 is one of the stress-responsive MAPK kinase kinase. ASK1 plays an important role in the response to reactive oxygen species, endoplasmic reticulum stress, and pro-inflammatory cytokines (Kawarazaki et al., 2014). It is involved in the pathogenesis of various diseases, including cancer, neurodegenerative diseases, infections, diabetes, and cardiovascular diseases (Liu et al., 2000). Nuclear factor kappa B (NF_kB) plays essential roles in activating immune responses to exogenous stimuli and indigenous stimulation. NF_kB proteins translocate into the nucleus to perform its functions during the stress condition (Thanos and Maniatis, 1995). NF_kB expression was found in almost all cell types and

tissues (Oeckinghaus and Ghosh, 2009). Furthermore, NF_kB is involved in the mechanism of many processes, including immune and inflammatory responses, stress responses, and regulation of cell proliferation and apoptosis (Oeckinghaus and Ghosh, 2009).

Glycolysis/gluconeogenesis promotes the conversion of glucose ($C_6H_{12}O_6$) into pyruvate ($CH_3COCOO^- + H^+$), releasing free energy to form the high-energy molecule ATP and reduced nicotinamide adenine dinucleotide (Lubert, 1995). The main precursors of Glycolysis/gluconeogenesis include lactic acid, pyruvic acid, amino acid, and glycerol. In mammals, the process of Glycolysis/gluconeogenesis mainly occurred in the liver. The Glycolysis/gluconeogenesis ability of the kidney is only 10% of that of the liver in a normal condition. However, the ability of Glycolysis/gluconeogenesis in the kidney will significantly enhance at a starved condition for a long time. HXK proteins play important roles in catalyzing hexose phosphorylation and sugar sensing and signaling. Sucrose must be cleaved in hexoses (glucose and fructose) during the sucrose metabolism (Siemens et al., 2011). The hexoses are phosphorylated by HXKs or fructokinases. The phosphorylated hexoses are then involved in metabolic processes (David-Schwartz et al., 2013). PPT1 plays essential roles in catalyzing the hydrolysis of lipid thioesters on *S*-acylated proteins, which has a lysosomal localization and function in non-neuronal cells (Lu et al., 1996; Verkruyse and Hofmann, 1996; Lu et al., 2002). Individuals lacking functional PPT1 present with progressive psychomotor decline within the first year of life, followed by loss of vision and profound seizures, before entering a persistent vegetative state that invariably ends in premature death (Santavuori et al., 1973; Hofmann et al., 2002).

The functions of $NF_kB\alpha$ were further analyzed by qPCR, *in situ* hybridization, RNAi, and histological observations, because $NF_kB\alpha$ was the most up-regulated gene in the double-side eyestalk ablation prawns. The previous studies and histological observation revealed that the hormones secreted by eyestalk have negative effects on testis development (Sroyraya et al., 2010; Chung et al., 2011; Guo et al., 2019). The histological observations after the ablation of eyestalk in *M. nipponense* revealed that the testis development became vigorous after eyestalk ablation, and RNA-Seq analysis predicted that Apoptosis played essential roles in the maintaining the normal testis development after eyestalk ablation through digesting the aged cells. $NF_kB\alpha$ is enriched in the metabolic pathway of Apoptosis, which has essential DEGs, differentially expressed in all of these three comparisons. Thus, the significant up-regulation of $NF_kB\alpha$ expression after eyestalk ablation indicated that $NF_kB\alpha$ was predicted as a strong candidate gene for the mechanism of male sexual development in *M. nipponense*. The previous studies reported that the activation and repression of the $NF_kB\alpha$ expression were involved in the pathogenesis of inflammatory diseases, such as adult respiratory distress syndrome (ARDS) and breast cancers (Blackwell and Christman, 1997; Biswas et al., 2001). $NF_kB\alpha$ was proven to be involved in the immune system of *Macrobrachium rosenbergii*, because the expressions of $NF_kB\alpha$ were up-regulated after the feed of scutellaria polysaccharide and soybean antigen protein (Yang et al., 2019). However, to the best of our knowledge, no study has reported the potential functions of $NF_kB\alpha$ in the mechanism of male sexual development in any species. The qPCR analysis in different mature tissues revealed that the highest expression of Mn-$NF_kB\alpha$ was observed in the testis, which was significantly higher than the other tested tissues and showed a significant difference with other tested tissues, indicating that Mn-$NF_kB\alpha$ may have potential functions during the testis development in *M. nipponense*. qPCR was also used to measure the Mn-$NF_kB\alpha$ expression in post-larval developmental stages of *M. nipponense*. The results revealed that the Mn-$NF_kB\alpha$ expression was gradually increased with the specimen development, and PL25♂ showed higher expression than that of PL25♀. The sensitive period of gonad differentiation and development of *M. nipponense* has been proven to be from PL7 to PL22 (Jin et al., 2016). Thus, Mn-$NF_kB\alpha$ was predicted to play essential roles in male sexual development in *M. nipponense*, combined with the qPCR analysis in different mature tissues and post-larval developmental stages. *In situ* hybridization revealed that signals were observed in spermatogonia and spermatocytes, indicating that Mn-$NF_kB\alpha$ played essential roles in the testis development in *M. nipponense*. No signal was directly observed in the androgenic gland cells, while strong signals were observed in the ejaculatory bulb surrounding the androgenic gland cells, indicating that Mn-$NF_kB\alpha$ has potential functions in maintaining the normal functions and structures of androgenic gland in *M. nipponense* (Jin et al., 2018, 2019). In different ovarian developmental stages, no signal was observed in O I and O V, while signals were observed in the nucleus, yolk granule, yolk granule, and cytoplasmic membrane in O II, O III, and O IV, indicating that Mn-$NF_kB\alpha$ promotes yolk accumulation in *M. nipponense* (Li et al., 2018). RNAi analysis revealed that the ds-RNA of Mn-$NF_kB\alpha$ can efficiently knockdown the expression of Mn-$NF_kB\alpha$ in *M. nipponense*. In addition, the expression of Mn-IAG was also decreased with the decrease of Mn-$NF_kB\alpha$, indicating that Mn-$NF_kB\alpha$ has a positive regulatory relationship with Mn-IAG. Thus, Mn-$NF_kB\alpha$ was involved in the male sexual development in *M. nipponense*, based on the importance of IAG in the male sexual development in crustacean species (Ventura et al., 2009, 2011, 2012). Histological observations after the treatment of Mn-$NF_kB\alpha$ dsRNA revealed that the number of sperms was decreased with the time of Mn-$NF_kB\alpha$ dsRNA treatment, indicating that Mn-$NF_kB\alpha$ has positive effects on testis development in *M. nipponense*.

In conclusion, histological observations revealed that eyestalk has negative effects on male sexual development in *M. nipponense*. A total of 1,039, 1,226, and 3,682 DEGs were identified between CG vs SS, SS vs DS, and CG vs DS, respectively, indicating that the ablation of double-side eyestalk has more regulatory roles on male sexual development in *M. nipponense*. Lysosome, Apoptosis, Glycolysis/Gluconeogenesis, and Insulin signaling pathway were the main enriched metabolic pathways in all of these three comparisons, and 10 important genes from these metabolic pathways were also selected. The functional analysis of $NF_kB\alpha$ by qPCR, RNAi, and histological observations revealed that $NF_kB\alpha$ has a positive regulatory effect on testis development in *M. nipponense*. This study identified the important functions of $NF_kB\alpha$ in male sexual development

in *M. nipponense*, providing new insights for the construction of the technique to regulate the testis development. Crisper9 techniques will be further used to knock out the gene expression of $NF_kB\alpha$ in *M. nipponense* and to identify whether $NF_kB\alpha$ is also an important gene in the mechanism of sex determination in *M. nipponense*, resulting in the sea reversal.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the NCBI repository, accession numbers: SRX9832767–SRX9832775.

ETHICS STATEMENT

The animal study was reviewed and approved by *Macrabrachium nipponense* the committee of Freshwater Fisheries Research Center and the Tai Lake Fishery Management Council. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

ShJ designed and wrote the manuscript. HF supervised the study. YH performed the eyestalk ablation and transcriptome profiling analysis. YF and YG revised the manuscript. HQ performed the qPCR analysis. WZ performed the *in situ* hybridization analysis. YX performed the RNAi analysis. YW performed the histological observations. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by grants from Central Public-interest Scientific Institution Basal Research Fund CAFS (2021JBFM02 and 2020TD36); the National Key R&D Program of China (2018YFD0900201 and 2018YFD0901303); Jiangsu Agricultural Industry Technology System; the China Agriculture Research System-48 (CARS-48); and the New Cultivar Breeding Major Project of Jiangsu Province (PZCZ201745).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.675928/full#supplementary-material

Supplementary Table 1 | Summary of BLASTx results for unigenes of testis *M. nipponense* transcriptome.

REFERENCES

Abu-Elheiga, L., Matzuk, M. M., Abo-Hashema, K. A., and Wakil, S. J. (2001). Continuous fatty acid oxidation and reduced fat storage in mice lacking acetyl-CoA carboxylase 2. *Science* 291, 2613–2616. doi: 10.1126/science.1056843

Almeida, E. A., Petersen, R. L., Andreatta, E. R., and Bainy, A. C. (2004). Effects of captivity and eyestalk ablation on antioxidant status of shrimps (*Farfantepenaeus paulensis*). *Aquaculture* 238, 523–528. doi: 10.1016/j.aquaculture.2004.04.010

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, L. (2001). Controlling the false discovery rate in behavior genetics research. *Behav. Brain. Res.* 125, 279–284. doi: 10.1016/s0166-4328(01)00297-2

Biswas, D. K., Dai, S. C., Cruz, A., Weiser, B., Graner, E., and Pardee, A. B. (2001). The nuclear factor kappa B (NF-κB): a potential therapeutic target for estrogen receptor negative breast cancers. *Proc. Natl. Acad. Sci. U.S.A.* 98, 10386–10391. doi: 10.1073/pnas.151257998

Blackwell, T. S., and Christman, J. W. (1997). The role of nuclear factor-kappa B in cytokine gene regulation. *Am. J. Resp. Cell. Mol.* 17, 3–9.

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

Cai, Y., and Shokita, S. (2006). Report on a collection of freshwater shrimps (Crustacea: Decapoda: Caridea) from the Philippines, with descriptions of four new species. *Raffles B. Zool.* 54, 245–270.

Cao, J. X. (2006). *The Molecular Mechanism in Male Reproductive Tract of the Prawn, Macrobrachium Rosenbergii*. Hangzhou: Zhejiang University.

Chung, J. S., Manor, R., and Sagi, A. (2011). Cloning of an insulin-like androgenic gland factor (IAG) from the blue crab, *Callinectes sapidus*: implications for eyestalk regulation of IAG expression. *Gen. Comp. Endocrinol.* 173, 4–10. doi: 10.1016/j.ygcen.2011.04.017

David-Schwartz, R., Weintraub, L., Vidavski, R., Zemach, H., Murakhovsky, L., Swartzberg, D., et al. (2013). The SLFRK4 promoter is active only during late stages of pollen and anther development. *Plant Sci.* 199, 61–70. doi: 10.1016/j.plantsci.2012.09.016

Diarte-Plata, G., Sainz-Hernández, J. C., Aguinaga-Cruz, J. A., Fierro-Coronado, J. A., Polanco-Torres, A., and Puente-Palazuelos, C. (2012). Eyestalk ablation procedures tominimize pain in the freshwater prawn *Macrobrachium Americanum*. *Appl. Anim. Behav. Sci.* 140, 172–178. doi: 10.1016/j.applanim.2012.06.002

Duve, C. D., and Wattiaux, R. (1966). Function of lysosomes. *Annu. Rev. Physiol.* 28, 435–492.

Fukushima, H., Wet, J. R., and O'Brien, J. S. (1985). Molecular cloning of a cDNA for human alpha-L-fucosidase. *Proc. Natl. Acad. Sci. U.S.A.* 82, 1262–1265. doi: 10.1073/pnas.82.4.1262

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652.

Guo, Q., Li, S., and Lv, X. (2019). Sex-biased CHHs and their putative receptor regulate the expression of IAG gene in the shrimp *Litopenaeus vannamei*. *Front. Physiol.* 10:1525.

Guo, Z. H. (2007). *Study on Proliferation and Differentiation of Spermatogenic Cells From Macrobrachium Nipponense in Vitro*. Baoding: Hebei University.

Hofmann, S. L., Atashband, A., Cho, S. K., Das, A. K., Gupta, P., Lu, J. Y., et al. (2002). Neuronal ceroid lipofuscinoses caused by defects in soluble lysosomal enzymes (CLN1 and CLN2). *Curr. Mol. Med.* 2, 423–437. doi: 10.2174/1566524023362294

Hopkins, P. M. (2012). The eyes have it: a brief history of crustacean neuroendocrinology. *Gen. Comp. Endocrinol.* 175, 357–366. doi: 10.1016/j.ygcen.2011.12.002

Hu, Y. N., Fu, H. T., Qiao, H., Sun, S. M., Zhang, W. Y., Jin, S. B., et al. (2018). Validation and evaluation of reference genes for Quantitative real-time PCR in *Macrobrachium nipponense*. *Int. J. Mol. Sci.* 19:2258. doi: 10.3390/ijms19082258

Jiang, F. W., Fu, H. T., Qiao, H., Zhang, W. Y., Jiang, S. F., Xiong, X. Y., et al. (2014). The RNA interference regularity of transformer-2 gene of oriental river prawn *Macrobrachium nipponense*. *Chin. Agricult. Sci. Bul.* 30, 32–37.

Jin, S. B., Fu, H. T., Jiang, S. F., Xiong, Y. W., Sun, S. M., Qiao, H., et al. (2018). Molecular cloning, expression, and in situ hybridization analysis of forkhead box protein L2 during development in *Macrobrachium nipponense*. *J. World Aquacult. Soc.* 49, 429–440. doi: 10.1111/jwas.12510

Jin, S. B., Fu, H. T., Sun, S. M., Jiang, S. F., Xiong, Y. W., Gong, Y. S., et al. (2017). Integrated analysis of microRNA and mRNA expression profiles at sex-differentiation sensitive period in oriental river prawn, *Macrobrachium nipponense*. *Sci. Rep.* 7:12011.

Jin, S. B., Fu, H. T., Zhou, Q., Sun, S. M., Jiang, S. F., Xiong, Y. W., et al. (2013a). Transcriptome analysis of androgenic gland for discovery of novel genes from the oriental river prawn, *Macrobrachium nipponense*, using Illumina Hiseq 2000. *PLoS One* 8:e76840. doi: 10.1371/journal.pone.0076840

Jin, S. B., Hu, Y. N., Fu, H. T., Jiang, S. F., Xiong, Y. W., Qiao, H., et al. (2019). Potential functions of Gem-associated protein 2-like isoform X1 in the oriental river prawn *Macrobrachium nipponense*: Cloning, qPCR, *in situ* hybridization, and RNAi analysis. *Int. J. Mol. Sci.* 20:3995. doi: 10.3390/ijms20163995

Jin, S. B., Wang, N., Qiao, H., Fu, H. T., Wu, Y., Gong, Y. S., et al. (2013b). Molecular cloning and expression of a full-length cDNA encoding crustacean hyperglycemic hormone (CHH) in oriental river pawn (*Macrobrachium nipponense*). *J. Fish. China.* 20, 82–92. doi: 10.3724/sp.j.1118.2013.00082

Jin, S. B., Zhang, Y., Guan, H. H., Fu, H. T., Jiang, S. F., Xiong, Y. W., et al. (2016). Histological observation of gonadal development during post-larva in oriental river prawn, *Macrobrachium nipponense*. *Chin. J. Fish.* 29, 11–16.

Jin, S., Hu, Y. N., Fu, H. T., Jiang, S. F., Xiong, Y. W., Qiao, H., et al. (2020). Analysis of testis metabolome and transcriptome from the oriental river prawn (*Macrobrachium nipponense*) in response to different temperatures and illumination times. *Comp. Biochem. Phys. D* 34:100662. doi: 10.1016/j.cbd.2020.100662

Johnstone, R. W., Ruefli, A. A., and Lowe, S. W. (2002). Apoptosis: a link between cancer genetics and chemotherapy. *Cell* 108, 153–164.

Kawarazaki, Y., Ichijo, H., and Naguro, I. (2014). Apoptosis signal-regulating kinase 1 as a therapeutic target. *Expert Opin. Ther. Tar.* 18, 651–664. doi: 10.1517/14728222.2014.896903

Li, F. J., Bai, H. K., Zhang, W. Y., Fu, H. T., Jiang, F. W., Liang, G. X., et al. (2015). Cloning of genomic sequences of three crustacean hyperglycemic hormone superfamily genes and elucidation of their roles of regulating insulin-like androgenic gland hormone gene. *Gene* 561, 68–75. doi: 10.1016/j.gene.2015.02.012

Li, F., Qiao, H., Fu, H. T., Sun, S. M., Zhang, W. Y., Jin, S. B., et al. (2018). Identification and characterization of opsin gene and its role in ovarian maturation in the oriental river prawn *Macrobrachium nipponense*. *Comp. Biochem. Physiol. B.* 218, 1–12. doi: 10.1016/j.cbpb.2017.12.016

Liu, H., Nishitoh, H., Ichijo, H., and Kyriakis, J. M. (2000). Activation of apoptosis signal-regulating kinase 1 (ASK1) by tumor necrosis factor receptor-associated factor 2 requires prior dissociation of the ASK1 inhibitor thioredoxin. *Mol. Cell. Biol.* 20, 2198–2208. doi: 10.1128/mcb.20.6.2198-2208.2000

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using realtime quantitative PCR and the 2-ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Lu, J. Y., Verkruyse, L. A., and Hofmann, S. L. (1996). Lipid thioesters derived from acylated proteins accumulate in infantile neuronal ceroid lipofuscinosis. Correction of the defect in lymphoblasts by recombinant palmitoyl protein thioesterase. *Proc. Natl. Acad. Sci. U.S.A.* 93, 10046–10050. doi: 10.1073/pnas.93.19.10046

Lu, J. Y., Verkruyse, L. A., and Hofmann, S. L. (2002). The effects of lysosomotropic agents on normal and INCL cells provide further evidence for the lysosomal nature of palmitoyl-protein thioesterase function. *Biochim. Biophys. Acta* 1583, 35–44. doi: 10.1016/s1388-1981(02)00158-0

Lubert, S. (ed.). (1995). "Glycolysis," in *Biochemistry*, 4th Edn. (New York, NY: W.H. Freeman and Company), 483–508.

Luzio, J. P., Pryor, P. R., and Bright, N. A. (2007). Lysosomes: fusion and function. *Nat. Rev. Mol. Cell. Biol.* 8, 622–632. doi: 10.1038/nrm2217

Ma, K. Y., Feng, J. B., Lin, J. Y., and Li, J. L. (2011). The complete mitochondrial genome of *Macrobrachium nipponense*. *Gene* 487, 160–165. doi: 10.1016/j.gene.2011.07.017

Ma, X. K., Liu, X. Z., Wen, H. S., Xu, Y. J., and Zhang, L. J. (2006). Histological observation on gonadal sex differentiation in *Cynoglossus semilaevis* Günther. *Mar. Fish. Res.* 27, 55–61.

Minoru, K., Michihiro, A., Susumu, G., Masahiro, H., Mika, H., Masumi, I., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic. Acids. Res.* 36, D480–D484.

Oeckinghaus, A., and Ghosh, S. (2009). The NF-$_k$B family of transcription factors and its regulation. *CSH Perspect. Biol.* 1:a000034. doi: 10.1101/cshperspect.a000034

Pamuru, R. R., Rosen, O., Manor, R., Chung, J. S., Zmora, N., Glazer, L., et al. (2012). Stimulation of molt by RNA interference of the molt inhibiting hormone in the crayfish *Cherax quadricarinatus*. *Gen. Comp. Endocrinol.* 178, 227–236. doi: 10.1016/j.ygcen.2012.05.007

Qiao, H., Fu, H. T., Jin, S. B., Wu, Y., Jiang, S. F., Gong, Y. S., et al. (2012). Constructing and random sequencing analysis of normalized cDNA library of testis tissue from oriental river prawn (*Macrobrachium nipponense*). *Comp. Biochem. Phys. D* 7, 268–276. doi: 10.1016/j.cbd.2012.04.003

Qiao, H., Jiang, F. W., Xiong, Y. W., Jiang, S. F., Fu, H. T., Li, F., et al. (2018). Characterization, expression patterns of molt-inhibiting hormone gene of *Macrobrachium nipponense* and its roles in molting and growth. *PLoS One* 13:e0198861. doi: 10.1371/journal.pone.0198861

Qiao, H., Xiong, Y., Zhang, W., Fu, H., Jiang, S., Sun, S. M., et al. (2015). Characterization, expression, and function analysis of gonad-inhibiting hormone in Oriental River prawn, *Macrobrachium nipponense* and its induced expression by temperature. *Comp. Biochem. Phys. A.* 185, 1–8. doi: 10.1016/j.cbpa.2015.03.005

Qiu, G. F., Du, N. S., and Lai, W. (1995). Studies on the male reproductive system of the freshwater prawn, *Macrobrachium nipponense*. *J. Shanghai Fish. Univ.* 4, 107–111.

Revathi, P., Vasanthi, L. A., Jeyanthi, S., Sankaralingam, S., Ramasubburayan, R., Prakash, S., et al. (2013). Impact of eyestalk ablation on the androgenic gland activity in the freshwater prawn *Macrobrachium rosenbergii* (De Man). *World* 5, 373–381.

Sainz-Hernández, J. C., Racotta, I. S., Dumas, S., and Hernández-López, J. (2008). Effect of unilateral and bilateral eyestalk ablation in Litopenaeus vannamei male and female on several metabolic and immunologic variables. *Aquaculture* 283, 188–193. doi: 10.1016/j.aquaculture.2008.07.002

Salma, U., Uddowla, M. H., Kim, M., Kim, J. M., Bo, K. K., Baek, H. J., et al. (2012). Five hepatopancreatic and one epidermal chitinases froma pandalid shrimp (Pandalopsis japonica): cloning and effects of eyestalk ablation on gene expression. *Comp. Biochem. Phys. B* 161, 197–207. doi: 10.1016/j.cbpb.2011.11.005

Salman, S. D., Page, T. J., Naser, M. D., and Yasser, A. G. (2006). The invasion of *Macrobrachium nipponense* (De Haan, 1849) (Caridea: Palaemonidae) into the southern Iraqi marshes. *Aquat. Invasions* 1, 109–115. doi: 10.3391/ai.2006.1.3.2

Santavuori, P., Haltia, M., Rapola, J., and Raitta, C. (1973). Infantile type of so-called neuronal ceroid-lipofuscinosis: 1. A clinical study of 15 patients. *J. Neurol. Sci.* 18, 257–267. doi: 10.1016/0022-510x(73)90075-0

Santos, E. A., Eduardo, L., Nery, M., Goncalves, A. A., and Keller, R. (1997). Evidence for the involvement of the crustacean hyperglycemic hormone in the regulation of lipid metabolism. *Physiol. Biochem. Zool.* 70, 415–420. doi: 10.1086/515846

ShangGuan, B. M., Liu, Z. Z., and Li, S. Q. (1991). Histological studies on ovarian development in *Scylla serrata*. *J. Fish. China.* 15, 96–103.

Shen, H., Zhou, X., Bai, A., Ren, X., and Zhang, Y. (2013). Ecdysone receptor gene from the freshwater prawn *Macrobrachium nipponense*: identification of different splice variants and sexually dimorphic expression, fluctuation of expression in the molt cycle and effect of eyestalk ablation. *Gen. Comp. Endocrinol.* 193, 86–94. doi: 10.1016/j.ygcen.2013.07.014

Siemens, J., GonzáLez, M. C., Wolf, S., Hofmann, C., Greiner, S., Du, Y. J., et al. (2011). Extracellular invertase is involved in the regulation of clubroot disease in *Arabidopsis thaliana*. *Mol. Plant Pathol.* 12, 247–262. doi: 10.1111/j.1364-3703.2010.00667.x

Sroyraya, M., Chotwiwatthanakunab, C., Stewartc, M. J., Soonklangd, N., Kornthong, N., Phoungpetchara, I., et al. (2010). Bilateral eyestalk ablation of the blue swimmer crab, *Portunus pelagicus*, produces hypertrophy of the androgenic gland and an increase of cells producing insulin-like androgenic gland hormone. *Tissue Cell* 42, 293–300. doi: 10.1016/j.tice.2010.07.003

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*. 4:41.

Thanos, D., and Maniatis, T. (1995). NF-KB: A lesson in family values. *Cell* 80, 529–532. doi: 10.1016/0092-8674(95)90506-5

Thompson, C. (1995). Apoptosis in the pathogenesis and treatment of disease. *Science* 267, 1456–1462. doi: 10.1126/science.7878464

Tiu, S. H. K., and Chan, S. M. (2007). The use of recombinant protein and RNA interference approaches to study the reproductive functions of a gonad-stimulating hormone from the shrimp *Metapenaeus ensis*. *FEBS J.* 274, 4385–4395. doi: 10.1111/j.1742-4658.2007.05968.x

Treerattrakool, S., Chartthai, C., Phromma-in, N., Panyim, S., and Udomkit, A. (2013). Silencing of gonad-inhibiting hormone gene expression in Penaeus monodon by feeding with GIH dsRNA-enriched Artemia. *Aquaculture* 404, 116–121. doi: 10.1016/j.aquaculture.2013.04.024

Treerattrakool, S., Panyim, S., and Udomkit, A. (2011). Induction of ovarian maturation and spawning in Penaeus monodon broodstock by double-stranded RNA. *Mar. Biotechnol.* 13, 163–169. doi: 10.1007/s10126-010-9276-0

Ventura, T., Manor, R., Aflalo, E. D., Rosen, O., and Sagi, A. (2012). Timing sexual differentiation: full functional sex reversal achieved through silencing of a single insulin-like gene in the prawn, *Macrobrachium rosenbergii*. *Biol. Reprod.* 86:90.

Ventura, T., Manor, R., Aflalo, E. D., Weil, S., and Sagi, A. (2009). Temporal silencing of an androgenic gland-specific insulin-like gene affecting phenotypical gender differences and spermatogenesis. *Endocrinology* 150, 1278–1286. doi: 10.1210/en.2008-0906

Ventura, T., Manor, R., Aflalo, E. D., Weil, S., Khalaila, I., Rosen, O., et al. (2011). Expression of an androgenic gland-specific insulin-like peptide during the course of prawn sexual and morphotypic differentiation. *ISRN Endocrinol.* 2011:476283.

Verkruyse, L. A., and Hofmann, S. L. (1996). Lysosomal targeting of palmitoyl protein thioesterase. *J. Biol. Chem.* 271, 15831–15836. doi: 10.1074/jbc.271.26.15831

Wang, Y. B., Jin, S. B., Fu, H. T., Qiao, H., Sun, S. M., Zhang, W. Y., et al. (2019). Identification and characterization of the DMRT11E gene in the oriental river prawn *Macrobrachium nipponense*. *Int. J. Mol. Sci.* 20, 1734. doi: 10.3390/ijms20071734

Winder, W. W., and Hardie, D. G. (1996). Inactivation of acetyl-CoA carboxylase and activation of AMP-activated protein kinase in muscle during exercise. *Am. J. Physiol.* 270(2 Pt 1), E299–E304.

Yang, J., Guo, Z., Cai, X., Hua, X., Liu, T., Kong, C., et al. (2019). Physiological, biochemical, and immune effects of dietary soybean antigen proteins in the giant river prawn (*Macrobrachium rosenbergii*). *J. Fish. Sci. China* 26:322. doi: 10.3724/sp.j.1118.2019.18092

Yang, W. X., Du, N. S., and Lai, W. (1999). Functional 646 relationship between spermatogenic cells and Sertoli cells during spermatogenesis of freshwater shrimp, *Macrobrachium nipponense*. *Acta Zool. Sin.* 45, 178–186.

Zhang, Y. P., Qiao, H., Zhang, W. Y., Sun, S. M., Jiang, S. F., Gong, Y. S., et al. (2013a). Molecular cloning and expression analysis two sex-lethal homolog genes during development in Oriental river prawn, *Macrobrachium nipponense*. *Genet. Mol. Res.* 12, 4698–4711. doi: 10.4238/2013.october.18.8

Zhang, Y. P., Fu, H. T., Qiao, H., Jin, S. B., Gong, Y. S., Jiang, S. F., et al. (2013b). cDNA cloning, characterization and expression analysis of a transformer-2 gene in the oriental river prawn, *Macrobrachium nipponense*. *J. World Aquacult. Soc.* 44, 338–349.

Zhang, Y. P., Jiang, S. F., Xiong, Y. W., Sun, S. M., Qiao, H., Jin, S. B., et al. (2013c). Molecular cloning and expression analysis of extra sex combs gene during development in *Macrobrachium nipponense*. *Turk. J. Fish. Aquat. Sc.* 13, 331–340.

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.

frontiers in Genetics

DATA REPORT
published: 18 August 2021
doi: 10.3389/fgene.2021.707546

Chromosome Genome Assembly and Annotation of the *Capitulum mitella* With PacBio and Hi-C Sequencing Data

*Duo Chen, Xuehai Zheng, Zhen Huang, Youqiang Chen, Ting Xue, Ke Li, Xiaozhen Rao and Gang Lin**

The Public Service Platform for Industrialization Development Technology of Marine Biological Medicine and Products of the State Oceanic Administration, Fujian Key Laboratory of Special Marine Bioresource Sustainable Utilization, Southern Institute of Oceanography, College of Life Sciences, Fujian Normal University, Fuzhou, China

Keywords: ***Capitulum mitella*, chromosome genome assembly, PacBio, Hi-C analysis, genome evolution analysis**

OPEN ACCESS

Edited by:
Qiang Lin,
Chinese Academy of Sciences, China

Reviewed by:
Xinhai Ye,
Zhejiang University, China
Yuanning Li,
Yale University, United States

****Correspondence:***
Gang Lin
lgffz@fjnu.edu.cn

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

Received: *10 May 2021*
Accepted: *12 July 2021*
Published: *18 August 2021*

Citation:
Chen D, Zheng X, Huang Z, Chen Y, Xue T, Li K, Rao X and Lin G (2021) Chromosome Genome Assembly and Annotation of the Capitulum mitella With PacBio and Hi-C Sequencing Data. Front. Genet. 12:707546. doi: 10.3389/fgene.2021.707546

INTRODUCTION

A pedunculate barnacle *Capitulum mitella* (Linnaeus, 1758) is a dominant intertidal cirripede (Lee et al., 2000). *C. mitella* distribute in the mid- and high-intertidal zone of tropical and subtropical coasts and is common in the East Sea and South Sea of China. *C. mitella* has a great commercial value in China, the petiole muscle of which has been a source of traditional delicious seafood (Yuan et al., 2016). The local market demand led to high-intensity catching activity, and this behavior caused significant decrease in quantity and habitat destruction in recent years. Artificial aquaculture of fishery resources will help to reduce overexploitation of natural stocks and meet market needs for fishery products.

In general, the barnacle life cycle consists of six nauplius instars and a nonfeeding cyprid instar. In the barnacle life history, the cyprid is a transitory phase between the pelagic and sessile lifestyle. The cyprid role is responsible for locating, exploring, and attaching to a suitable substratum; subsequently, a complex metamorphosis and permanent settlement happen (Lagersson and Høeg, 2002). Cypris attachment and metamorphosis are also known as "cypris settlement" (Clare and Matsumura, 2000; Franco et al., 2016). The cypris settlement is a crucial event for the survival and development of both adults and subsequent generations. Within the last three decades, the basic biology of *C. mitella* such as reproductive characteristics, sperm ultrastructure, and larval culture conditions was investigated in our lab. *C. mitella* cyprids have been cultured on a relatively large scale. However, they cannot attach and metamorphose into juveniles in an artificial environment, so they would die eventually and interrupt their life history.

In recent years, the cypris morphology of *C. mitella* has been depicted minutely using SEM (Rao and Lin, 2014). After we had successfully induced metamorphosis of the *C. mitella* cyprids into juveniles, we described a time line and specific morphological changes during the metamorphosis under microscopy and SEM (Lin and Rao, 2017). The impacts of external (i.e., water temperature and salinity) and internal (i.e., cyprid age) factors on the metamorphosis of *C. mitella* were researched (Rao and Lin, 2020). These efforts are possible for completing the life history in an artificial environment and aquaculture of *C. mitella*. Not only is the artificial aquaculture for *C. mitella* a production activity but also it helps to reduce the stress on overexploited populations.

Barnacles are frequently dominant marine foulings and widely distributed around the world that play crucial roles in marine ecology (Maréchal and Hellio, 2011). Recently, there are a few transcriptome researches on barnacles, such as a model animal *Amphibalanus amphitrite* (Chen et al., 2011; Yan et al., 2012; Chandramouli et al., 2015; Sarah et al., 2016), a giant barnacle

Megabalanus volcano (Yan et al., 2017) and a stalked barnacle *Neolepas marisindica* (Ryu et al., 2019). These efforts showed that the transcriptome databases could greatly promote the research on molecular mechanisms of barnacle settlement. Genome sequencing has been performed for an acorn barnacle *Amphibalanus amphitrite*, but there is still no genome sequencing information for a stalked barnacle. Numerous reports on *C. mitella* were mainly related to mitochondrial DNA sequencing and quantitative proteomics analyses (Song and Yoon, 2013; Yoon et al., 2013; Yuan et al., 2016; Tian et al., 2020). So far, no genome literature on the *C. mitella* is available. Whereas a lack of genome information on the barnacles has obstructed further inquiry on the molecular mechanisms with development, attachment, metamorphosis, genetic evolution, and so on, so does on *C. mitella*.

Accordingly, in this study, we aimed to sequence and assemble the genome of *C. mitella*. The baseline data obtained from this study will be very useful in taxonomical identification, phylogenetic analysis, larval settlement mechanisms, artificial breeding and aquaculture, and species protection for *C. mitella*.

TABLE 1 | Assembly statistics for *Capitulum mitella*.

Items	Canu		Hi-C	
	Contig_len (Mb)	Contig_ number	Scaffold_len (Mb)	Scaffold_ number
Total	463.09	1,462	482.98	269
Max	11.04	–	45.43	–
Number ≥2 kb	–	1,462	–	217
N50	3.14	–	31.03	–

MATERIALS AND METHODS

Sample Preparation and Genome Sequencing

C. mitella specimens were sampled from the intertidal zone of Dinghai (26.2487°N, 119.7987°E), Fujian Province, China, and immediately frozen in liquid nitrogen. Total DNA was

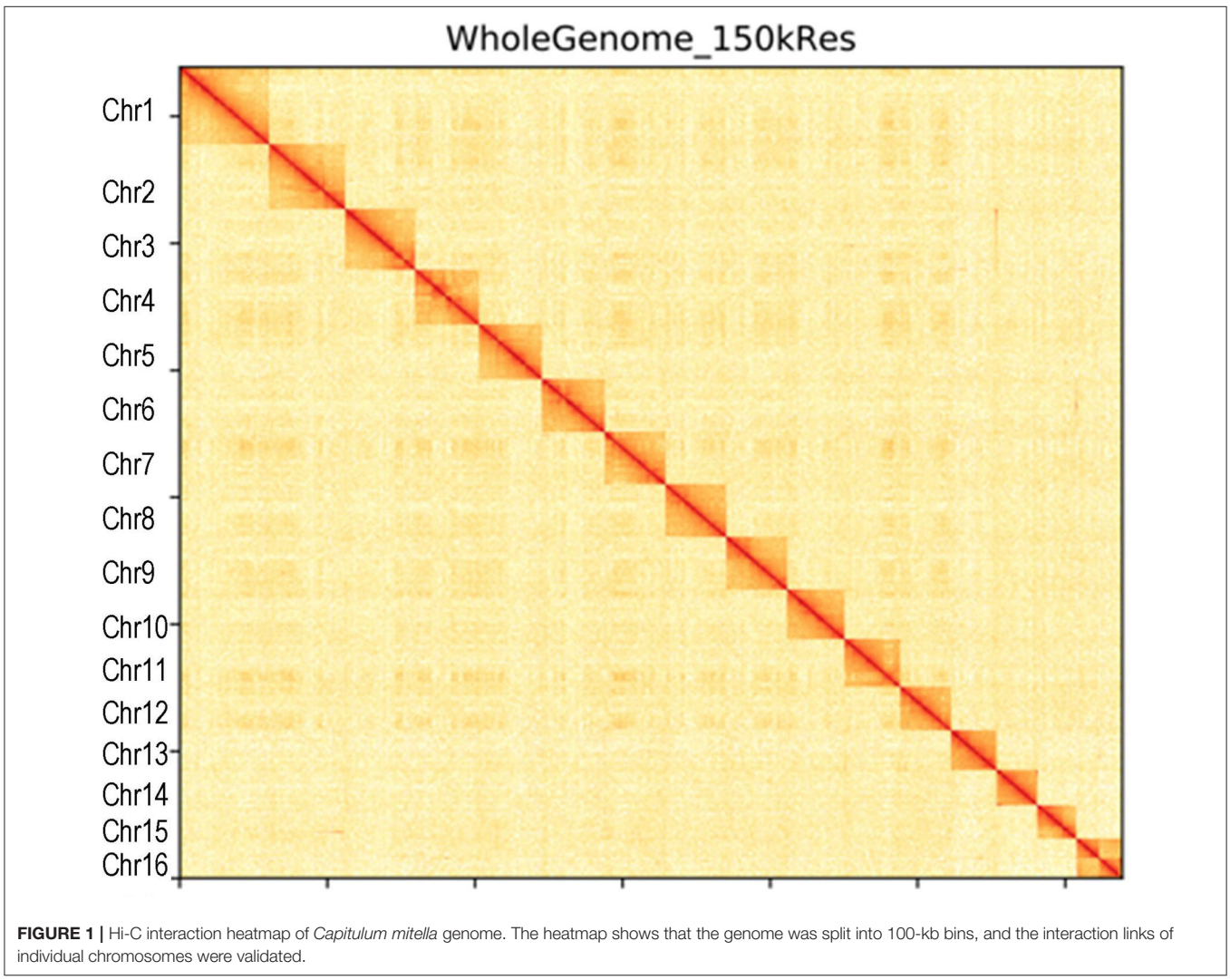

FIGURE 1 | Hi-C interaction heatmap of *Capitulum mitella* genome. The heatmap shows that the genome was split into 100-kb bins, and the interaction links of individual chromosomes were validated.

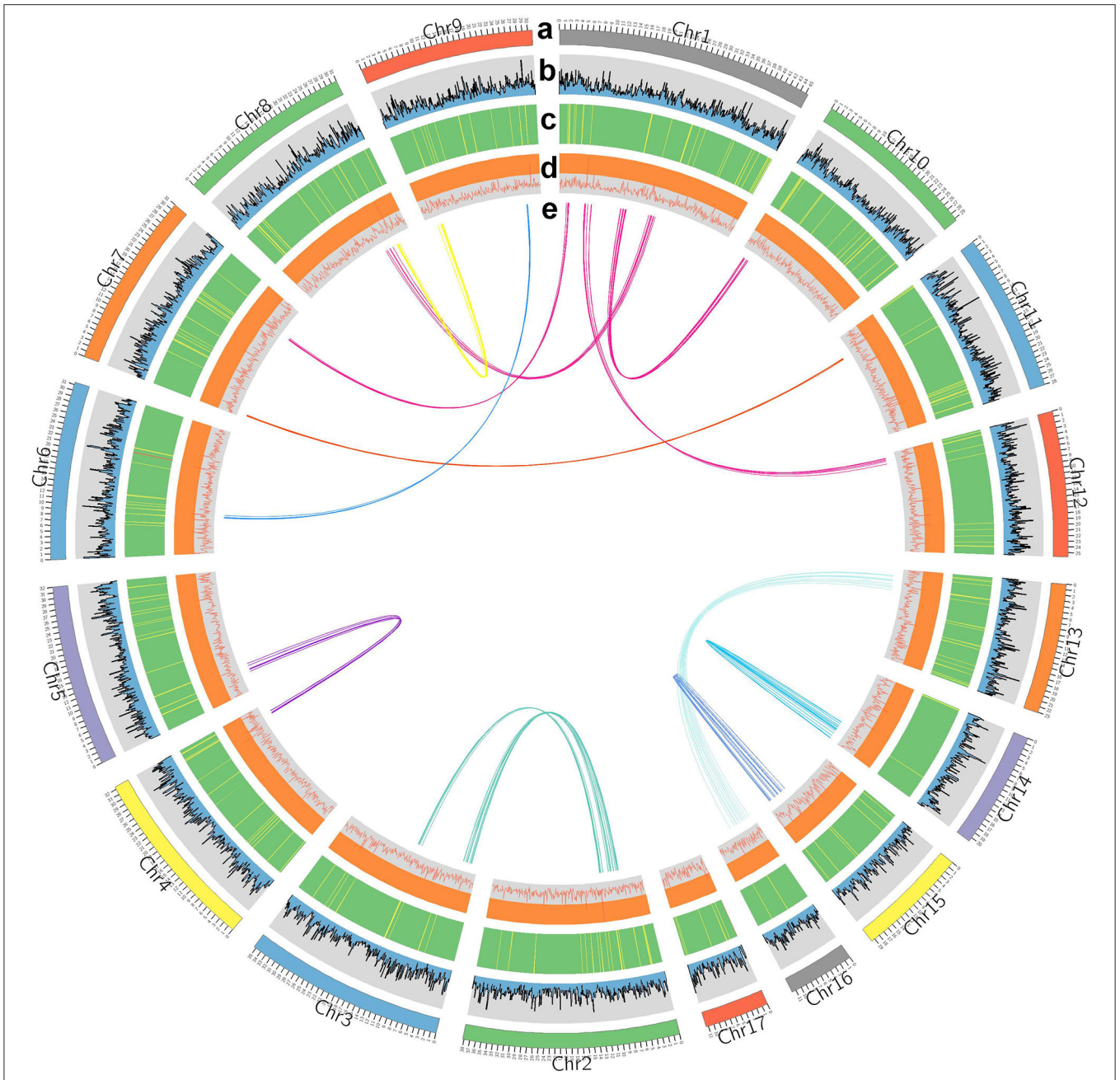

FIGURE 2 | Characterization of *Capitulum mitella*. Genome elements are arranged in the Circos plot. **(a)** Eight assembled chromosomes of the genome. **(b)** The distribution of GC content. **(c)** Distribution of gene density (green). **(d)** Statistics for coverage of TEs. **(e)** Circos diagram of major interchromosomal relationships in the *C. mitella* genome. Each line represents a syntenic block; block size = 3 kb.

isolated from fresh muscle samples with the DNeasy® Tissue Kit (Qiagen, Hilden, Germany) according to the DNeasy® Protocol for animal tissues. Qubit 3.0 (Thermo Fisher Scientific, Inc., Carlsbad, CA, USA) was used to assess the DNA concentration and quality before downstream sequencing. Approximately 5 μg of high-quality genomic DNA was sheared and size-selected (∼40 kb) for PacBio library construction followed by sequencing on the PacBio Sequel II platform (PacBio Biosciences, Menlo Park, CA, USA). Sequence libraries with inserts of average 350 bp were constructed for Illumina paired-end (PE) sequencing according to the manufacturer's instructions, and sequencing was subsequently performed on Illumina HiSeq X Ten platform (Illumina, San Diego, CA, USA).

TABLE 2 | Assessment of the completeness of the *Capitulum mitella* genome assembly by BUSCO.

Type	Number	Percent (%)
Complete BUSCOs (C)	957	94.4
Complete and single-copy BUSCOs (S)	951	90.3
Complete and duplicated BUSCOs (D)	42	4.1
Fragmented BUSCOs (F)	15	1.5
Missing BUSCOs (M)	41	4.1
Total BUSCO groups searched	1,013	100

TABLE 3 | General function annotation results statistics.

Annotation statistics for nuclear genome	Number	Percent (%)
Total protein	21,899	
EggNOG	16,224	74.08
GO	9,312	42.52
COG	14,204	64.86
KEGG	5,368	24.51
At least in one database	18,176	82.99

COG, Cluster of Orthologous Groups; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Genome Assembly

Canu (V2.0, https://github.com/marbl/canu/) (Koren et al., 2017) was used to perform the genome *de novo* assembly based on the PacBio long reads. After the initial assembly, primary contigs were corrected for three cycles with the Illumina reads using Pilon (V1.24, https://github.com/broadinstitute/pilon) (Belton et al., 2012). We used fresh muscle tissue from the same individual as in genome sequencing for the Hi-C library construction and sequencing. The Hi-C library was generated using Mbo I restriction enzyme following *in situ* ligation protocols (Peng et al., 2021). In brief, fresh muscle tissue was sampled and frozen in liquid nitrogen. Tissues were cross-linked with fresh formaldehyde and quenched with glycine. The total DNA was extracted from the nuclei in the lysis buffer [10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% NP40, and complete protease inhibitors (Roche)]. The purified DNA was digested with Mbo I and was labeled by Biotin-14-dATP (Thermo Fisher Scientific) and then ligated by T4 DNA Ligase. After ligation, the cross-linking was incubated with proteinase K overnight; the ligated DNA was sheared into 200–600-bp fragments and then blunt-end repaired and A-tailed, followed by purification through biotin-streptavidin-mediated pull-down. Finally, the Hi-C libraries were quantified and sequenced on the Illumina HiSeq X Ten platform using a PE-150 module. The Hi-C sequencing data were then aligned to the scaffolds using BWA-MEM alignment algorithm (V1.2.2, https://github.com/lh3/bwa). Two ends of reads were independently aligned to the genome and only selected the read pairs for which both ends were uniquely aligned to the genome. The hiclib software (V0.5.7, http://github.com/hiclib) and a previously reported method (Gong et al., 2018) were applied to filter the Hi-C reads, and the interaction frequency was quantified and normalized among contigs. Lachesis (Burton et al., 2013) with default parameters was then applied to cluster contigs with the agglomerative hierarchical clustering method using the interaction matrix between sequences. To construct a high-quality assembly, we sequenced the genome by utilizing a combination of Illumina (∼78.52 Gb of raw data), PacBio single-molecule real-time (SMRT) sequencing (∼60.23 Gb of raw data), and Hi-C chromosome-scale scaffolding (∼44.40 Gb of raw data) (Belton et al., 2012). The raw reads were filtered out by certain low-quality scrap reads and/or by minimum CLR length. After read quality filtering by PacBio Quality checking tool (http://github.com/PacificBiosciences/), a total of 60.17 Gb of long-read sequence data were produced, comprising 4,342,961 subreads with a read N50 of 19.979 Kb. This initial genome assembly was 463.09 Mb in length with a contig N50 of 3.14 Mb (**Table 1**; **Figures 1**, **2**). A total of 259,360,062 reads (87.63 Gb) with a Q30 of 93.23% were generated for Hi-C analysis. Finally, 16 chromosomes anchored by the contigs were generated with the guidance of HiC reads; this result consists of the karyotype analysis (**Supplementary Figure 1**). This chromosome-level assembly has an average length of 28.11 Mb, with the shortest chromosome of 19.66 Mb (Chr16) and the longest one of 45.43 Mb (Chr1) (**Figure 2**). The scaffold N50 reached 3.14–31.03 Mb (**Table 1**), providing a chromosome-level genome assembly for *C. mitella*. The completeness of this assembled genome was assessed based on BUSCO analysis, which revealed that nearly 94.4% of the Arthropoda orthologs were included in the assembled (**Table 2**). These results indicated a high-quality assembly of *C. mitella* genome.

Genome Annotation

The assembled *C. mitella* genome was homology-based annotated using a collection of protein-coding genes and RNA sequencing data from nine animal species: *Apis mellifera*, *Caenorhabditis elegans*, *Crassostrea gigas*, *Daphnia pulex*, *Drosophila melanogaster*, *Mizuhopecten yessoensis*, *Limulus polyphemus*, *Penaeus vannamei*, and *Plutella xylostella*. *D. pulex* and *P. vannamei* belong to subphylum Crustacea. *D. melanogaster*, *P. xylostella*, *and A. mellifera* belong to the subphylum Hexapoda, which is closely related to *C. mitella. M. yessoensis* and *C. gigas* belong to phylum Mollusca, and their ontogeny is similar to that of *C. mitella* through larval metamorphosis and attachment. *De novo* gene predictions were carried out by the AUGUSTUS package (V2.0) provided by BRAKER (V2.1.4) (Hoff et al., 2015). The MAKER pipeline was employed to identify the 21,899 protein-coding genes (Holt and Yandell, 2011), integrating protein sequences and transcript genes from the *de novo* assembly of *C. mitella* transcriptome data. All protein-coding genes were functionally annotated BLASTP (V2.2.3) (Anoop and Agostinho, 2015) against the public protein sequence databases EggNOG, Gene Ontology (GO), Cluster of Orthologous Groups (COG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) with an E-value $\leq$1e−5. We functionally annotated 14,204, 9,312, 14,204, and 5,368 genes to EggNOG, GO, COG, and KEGG, respectively (**Table 3**). Small RNAs (sRNAs) and noncoding RNAs (ncRNAs) were predicted by Rfam and miRNA databases using tRNAsan-SE (V2.0) (Bernard

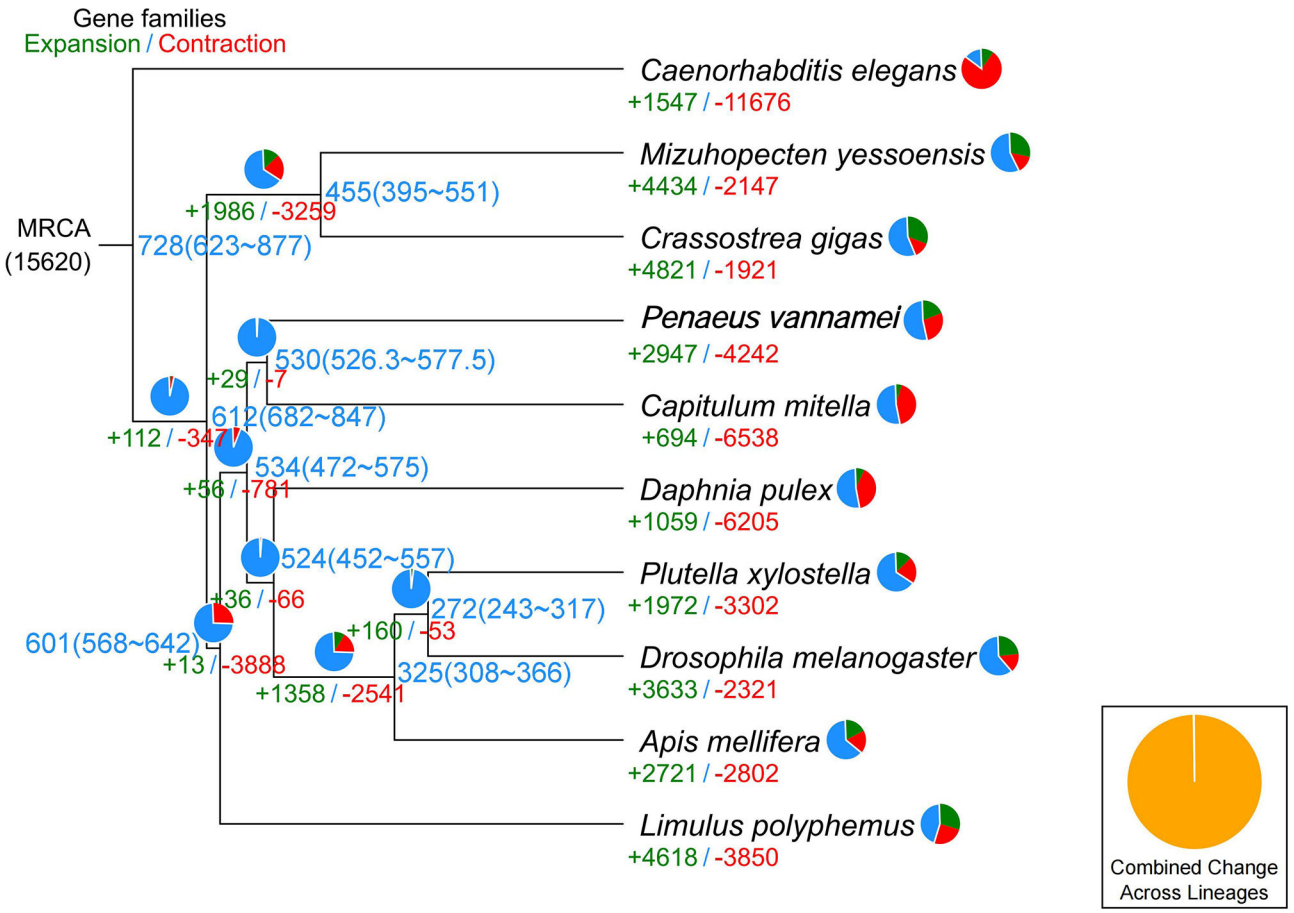

FIGURE 3 | Phylogenetic tree showing the number of gene families displaying expansion (green) and contraction (red) among 10 animal species. The pie charts show the proportions of expanded (green), contracted (red), and conserved (blue) gene families among all gene families. The estimated divergence time (million years ago) is displayed in the phylogenetic tree in black. MRCA, most recent common ancestor.

et al., 2014) and BLASTN (V2.2.3), respectively. Some types of ncRNAs consisting of microRNAs (miRNAs) and small nuclear RNAs (snRNAs) were identified using INFERNAL (V.1.1.4) by searching the Rfam database (http://infernal.janelia.org/). In total, 136 miRNAs, 748 tRNAs, 56 rRNAs, and 87 snRNAs were identified in *C. mitella* genome (**Supplementary Table 1**). Tandem Repeats Finder (V4.09.1), LTR_FINDER (V1.07), and RepeatMasker (V4.0) (González and Deyholos, 2012) were combined to predict repeat sequences in the *C. mitella* genome. In the *C. mitella* genome, repetitive sequences accounted for 3.53%, and long terminal repeat (LTR) retrotransposons accounted for 0.28% of the genome, including 0.18% Ty3/gypsy, 0.03% Ty1/copia, and 0.06% other (**Supplementary Table 2**).

Genome Evolution and Gene Family Analysis

In the current study, in order to investigate the evolution relationships among *C. mitella* and nine other species, we also perform phylogenomic analysis based on the amino acid sequences. These 10 species were *A. mellifera*, *C. mitella*, *C. elegans*, *C. gigas*, *D. melanogaster*, *D. pulex*, *M. yessoensis*, *L. polyphemus*, *P. vannamei*, and *P. xylostella*. Their protein sequences were acquired from the NCBI database, and single-copy genes were obtained by OrthoFinder (V2.27) (Emms and Kelly, 2015). Protein sequences of single-copy gene orthologs were aligned using MUSCLE software (V3.8.425). The alignments of Coding DNA Sequences (CDSs) that were guided by protein alignments were concatenated into the superalignment of nucleotide sequences. The phylogenetic relationships were then constructed using the maximum likelihood (ML) methods. On the basis of the protein alignments, the CDSs were aligned and then concatenated into a superalignment matrix for each family. The cluster size differences between the ancestor and each species were compared to analyze the expansion and contraction of the gene families using CAFE (V2.1) (Koichiro et al., 2007). The phylogenetic tree showed that *C. mitella* and *P. vannamei* phylogenetically were clustered into an independent ∼530

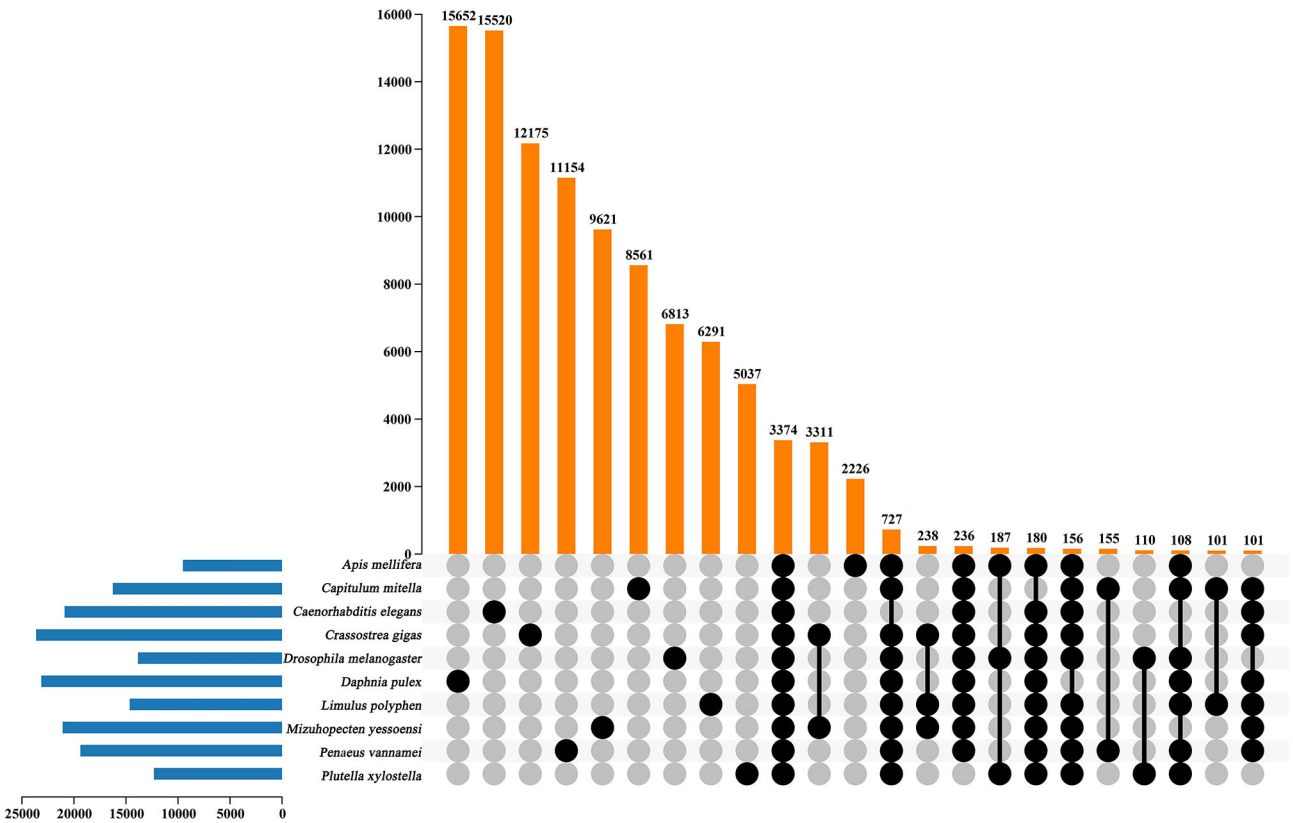

FIGURE 4 | UpSet plot of the intersection of gene families in *A. mellifera, C.mitella, C. elegans, C. gigas, D.melanogaster, D. pulex, M. yessoenisis, L. polyphemus, P. vannamei and P. xylostella*. The numbers of gene families (clusters) are indicated for each species and species intersection.

million years ago (Mya), which was earlier than those of Insecta (325 Mya) (**Figure 3**). Comparative genomic analyses were performed among abovementioned 10 animal species, and we detected 15,620 families of homologous genes, among which 7,232 gene families were identified in *C. mitella*, including 694 and 6,538 gene families showing contraction and expansion, respectively (**Figure 3**). GO analysis revealed that the 694 expanded orthogroups were involved in the metabolic process, developmental process, stimulus response, transporter activity, and catalytic activity (**Supplementary Table 3**). The analysis of KEGG pathways manifested that most of the 694 expanded genes were annotated to the signaling molecules and interaction, cell growth and death, glycine, environmental adaptation, and serine and threonine metabolism (**Supplementary Table 4**). The *C. mitella* is a common species in middle tidal areas of intertidal zones, distributed in rock surfaces or crevices. Therefore, *C. mitella* have evolved a complex mechanism to adapt to dramatically changing environments such as large temperature differences and desiccation. So, the expanded orthogroups in *C. mitella* are related to the metabolic process, stimulus process, environmental adaptation, and so on. By contrast, the group of contracted gene families was related to lipid metabolism, phenylalanine metabolism, signal transduction, immune system, environmental adaptation, and cell growth and death (**Supplementary Table 5**). The GO terms annotation of the contracted genes showed that they were most related to the stimulus response, biological regulation, catalytic activity, growth, reproductive process, signaling, and developmental process (**Supplementary Table 6**). The contracted gene families in this process related to the development in the insect, suggesting that *C. mitella* and the other compared species share the core signal regulation network in the developmental process. The comparison of *A. mellifera*, *C. mitella*, *C. elegans*, *C. gigas*, *D. melanogaster*, *D. pulex*, *M. yessoensis*, *L. polyphemus*, *P. vannamei*, and *P. xylostella* revealed that 3,374 (28.26%) of the 11,935 *C. mitella* gene families were shared by the other nine species, whereas 8,561 gene families were unique to *C. mitella* (**Figure 4**). GO functional analysis revealed that these 8,561 unique families were enriched in the immune system process, developmental process, signaling, aging, stimulus response, and calcium signaling pathway (**Supplementary Table 7**), while KEGG annotation presented that most of the 8,561 unique families were gathered in signal transduction and cysteine and methionine metabolism (**Supplementary Table 8**).

DATA AVAILABILITY STATEMENT

The whole genome sequence data reported in this paper available in the Genome Warehouse in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, under accession number GWHBAVK00000000.1 that is publicly accessible at https://bigd.big.ac.cn/gwh. The complete sequences generated in this study was deposited to the NCBI, under the accession PRJNA725059.

AUTHOR CONTRIBUTIONS

DC, XZ, ZH, and TX contributed to genome assembly and annotations. TX, ZH, and GL contributed to manuscript preparation. KL, ZH, GL, XR, and YC contributed to sampling and sequencing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the cooperation project on production and education of the University of Fujian province, China (2018N5007).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.707546/full#supplementary-material

Supplementary Figure 1 | The karyotype of *C. mitella*.

Supplementary Table 1 | Repetitive element annotations in *C. mitella*.

Supplementary Table 2 | Repetitive element annotations.

Supplementary Table 3 | GO enrichment of genes in the expansion families in *C. mitella* genome.

Supplementary Table 4 | KEGG enrichment of genes in the expansion families in *C. mitella* genome.

Supplementary Table 5 | KEGG enrichment of genes in the contraction families in *C. mitella* genome.

Supplementary Table 6 | GO enrichment of genes in the contraction families in *C. mitella* genome.

Supplementary Table 7 | GO enrichment of genes that were specifically identified in *C. mitella* genome.

Supplementary Table 8 | KEGG enrichment of genes that were specifically identified in *C. mitella* genome.

REFERENCES

Anoop, G. A., and Agostinho, A. (2015). Whole genome sequencing of the symbiont *pseudovibrio* sp. from the intertidal marine sponge *Polymastia penicillus* revealed a gene repertoire for host-switching permissive lifestyle. *Genome Biol. Evol.* 7, 3022–3032. doi: 10.1093/gbe/evv199

Belton, J. M., Mccord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., and Dekker, J. (2012). Hi-c: a comprehensive technique to capture the conformation of genomes. *Methods* 58, 268–276. doi: 10.1016/j.ymeth.2012.05.001

Bernard, B., Zofia, N., Le, B. A., and Bénédicte, C. (2014). Computational prediction and experimental validation of micrornas in the brown alga *Ectocarpus siliculosus*. *Nucleic Acids Res.* 42, 417–429. doi: 10.1093/nar/gkt856

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., and Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi: 10.1038/nbt.2727

Chandramouli, K. H., Al-Aqeel, S., Ryu, T., Zhang, H., Serinda, L., Ghosheh, Y., et al. (2015). Transcriptome and proteome dynamics in larvae of the barnacle *Balanus amphitrite* from the red sea. *BMC Genom.* 16:1063. doi: 10.1186/s12864-015-2262-1

Chen, Z. F., Kiyotaka, M., Hao, W., Arellano, S. M., Yan, X., Intikhab, A., et al. (2011). Toward an understanding of the molecular mechanisms of barnacle larval settlement: a comparative transcriptomic approach. *PLoS ONE.* 6:e22913. doi: 10.1371/journal.pone.0022913

Clare, A. S., and Matsumura, K. (2000). Nature and perception of barnacle settlement pheromones. *Biofouling* 15, 57–71. doi: 10.1080/08927010009386298

Emms, D. M., and Kelly, S. (2015). Orthofinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16, 1–14. doi: 10.1186/s13059-015-0721-2

Franco, S. C., Aldred, N., Cruz, T., and Clare, A. S. (2016). Modulation of gregarious settlement of the stalked barnacle, *Pollicipes pollicipes*: a laboratory study. *Sci. Mar.* 80, 217–228. doi: 10.3989/scimar.04342.01A

Gong, G., Dan, C., Xiao, S., Guo, W., Huang, P.,Xiong, Y., et al. (2018). Chromosomal-level assembly of yellow catfish genome using third-generation dna sequencing and hi-c analysis. *GigaScience* 7:giy120. doi: 10.1093/gigascience/giy120

González, L. G., and Deyholos, M. K. (2012). Identification, characterization and distribution of transposable elements in the flax (*Linum usitatissimum* L.) genome. *BMC Genom.* 13:644. doi: 10.1186/1471-2164-13-644

Hoff, K. J., Simone, L., Alexandre, L., Mark, B., and Mario, S. (2015). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 5, 767–769. doi: 10.1093/bioinformatics/btv661

Holt, C., and Yandell, M. (2011). Maker2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* 12:491. doi: 10.1186/1471-2105-12-491

Koichiro, T., Joel, D., Masatoshi, N., and Sudhir, K. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 8, 1596–1599. doi: 10.1093/molbev/msm092

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., and Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116

Lagersson, N. C., and Høeg, J. T. (2002). Settlement behavior and antennulary biomechanics in cypris larvae of *Balanus amphitrite* (crustacea: thecostraca: cirripedia). *Mar Biol.*141, 513–526. doi: 10.1007/s00227-002-0854-1

Lee, C., Shim, J. M., and Kim., C. H. (2000). Larval development of *Capitulum mitella* (Cirripedia: Pedunculata) reared in the laboratory. *J. Mar. Biol. Assoc. UK* 80, 457–464. doi: 10.1017/S0025315400002150

Lin, G., and Rao, X. (2017). Metamorphosis of the pedunculate barnacle *Capitulum mitella* Linnaeus, 1758 (Cirripedia: Scalpelliformes). *J. Mar. Biol. Assoc. UK.* 97, 1643–1650. doi: 10.1017/S0025315416001089

Maréchal, J. P., and Hellio, C. (2011). Antifouling activity against barnacle cypris larvae: do target species matter (*Amphibalanus amphitrite* versus *Semibalanus balanoides*). *Int. Biodeter. Biodegr.* 65, 92–101. doi: 10.1016/j.ibiod.2010.10.002

Peng, Y., Li, H., Liu, Zh., Zhang, C., Li, K.,Gong,Y., et al. (2021). Chromosome-level genome assembly of the Arctic fox (*Vulpes lagopus*) using PacBio sequencing and Hi-C technology. *Mol. Ecol. Resour.* 1, 1–16. doi: 10.1111/1755-0998.13397

Rao, X., and Lin, G. (2014). Scanning electron microscopy of the cypris larvae of *Capitulum mitella* (Cirripedia: Thoracica: Scalpellomorpha). *J. Mar. Biol. Assoc. UK* 94, 361–368. doi: 10.1017/S0025315413001173

Rao, X., and Lin, G. (2020). Effects of age, salinity and temperature on the metamorphosis and survival of *Capitulum mitella* cyprids (Cirripedia:

Thoracica: Scalpellomorpha). *J. Mar. Biol. Assoc. UK*. 100, 55–62. doi: 10.1017/S0025315419001152

Ryu, T., Woo, S., and Lee, N. (2019). The first reference transcriptome assembly of the stalked barnacle, *Neolepas marisindica,* from the onnuri vent field on the central indian ridge. *Mar. Genom*. 48, 100679–100682. doi: 10.1016/j.margen.2019.04.004

Sarah, A. A., Taewoo, R., Zhang, H., Chandramouli, K. H., and Timothy, R. (2016). Transcriptome and proteome studies reveal candidate attachment genes during the development of the barnacle *Amphibalanus amphitrite*. *Front. Mar. Sci.* 3:171. doi: 10.3389/fmars.2016.00171

Song, Y. J., and Yoon, J. M. (2013). Genetic differences of three *Pollicipesmitella*populations identified by pcr analysis. *J. Reprod. Dev*. 17, 199–205. doi: 10.12717/DR.2013.17.3.199

Tian, M., Chen, P., Song, J., He, F., and Shen, X. (2020). The first mitochondrial genome of *Capitulum mitella* (Crustacea: Cirripedia) from China: revealed the phylogenetic relationship within Thoracica. *Mitochondrial DNA B* 5, 2573–2575. doi: 10.1080/23802359.2020.1781564

Yan, G., Zhang, G., Huang, J., Yi, L., Sun, J., and Zeng, C., et al. (2017). Comparative transcriptomic analysis reveals candidate genes and pathways involved in larval settlement of the barnacle Megabalanus volcano. *Int. J. Mol. Sci.* 18, 2253–2267. doi: 10.3390/ijms181 12253

Yan, X. C., Chen, Z. F., Jin, S., Matsumura, K., and Wu, R. (2012). Transcriptomic analysis of neuropeptides and peptide hormones in the barnacle *Balanus amphitrite*: Evidence of roles in larval settlement. *PLoS ONE*. 7:e46513. doi: 10.1371/journal.pone.0046513

Yoon, M., Jung, J. Y., and Dong, S. K. (2013). Genetic diversity and gene flow patterns in *Pollicipesmitella* in korea inferred from mitochondrial dna sequence analysis. *Fish Aquat. Sci*. 16, 243–251. doi: 10.5657/FAS.2013.0243

Yuan, T. P., Huang, Y. P., Miao, S. Y., Li, L., and Yan, Y. (2016). Genetic diversity and population structure of *Capitulum mitella* (Cirripedia: Pedunculata) in China inferred from mitochondrial DNA sequences. *Biochem. Syst. Ecol*. 67, 22–28. doi: 10.1016/j.bse.2016.05.016

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 in Genetics

ORIGINAL RESEARCH
published: 19 August 2021
doi: 10.3389/fgene.2021.717764

Genome-Wide SNP Discovery and Population Genetic Analysis of *Mesocentrotus nudus* in China Seas

*Quanchao Wang[1,2], Ying Liu[1,3], Lang Yan[1], Linlin Chen[1,2] and Baoquan Li[1,2]**

[1] Key Laboratory of Coastal Biology and Bioresource Utilization, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, China, [2] Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China, [3] University of Chinese Academy of Sciences, Beijing, China

Mesocentrotus nudus is an important commercially aquatic species because of its high edible and medicinal values. However, wild stocks have dramatically decreased in recent decades. Understanding the population structure and genetic diversity can provide vital information for genetic conservation and improvement. In the present study, the genotyping-by-sequencing (GBS) approach was adopted to identify the genome-wide single-nucleotide polymorphisms (SNPs) from a collection of 80 individuals consisting of five geographical populations (16 individuals from each population), covering the natural habitats of *M. nudus* in China seas. An average of 0.96-Gb clean reads per sample were sequenced, and a total of 51,738 biallelic SNPs were identified. Based on these SNPs, diversity index analysis showed that all populations have a similar pattern with positive F_{is} (0.136) and low Ne (724.3). Low genetic differentiation and high genetic connectivity among five geographical populations were detected by pairwise F_{st}, principal component analysis (PCA), admixture, and phylogenetic analysis. Besides, two YWL individuals originating from an isolated ancestor may imply that there is a genetically differentiated population in the adjacent sea. Overall, the results showed that GBS is an effective method to detect genome-wide SNPs for *M. nudus* and suggested that the protective measures and the investigation with larger spatial scale and sample size for *M. nudus* should be carried out in the future.

Keywords: *Mesocentrotus nudus*, single-nucleotide polymorphism discovery, genetic diversity, population structure, China seas

OPEN ACCESS

Edited by:
Qiang Lin,
South China Sea Institute of Oceanology, Chinese Academy of Sciences, China

Reviewed by:
Tainã Figueiredo Cardoso, Brazilian Agricultural Research Corporation (EMBRAPA), Brazil
Zhiyi Bai, Shanghai Ocean University, China

***Correspondence:**
Baoquan Li
bqli@yic.ac.cn

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

Received: *31 May 2021*
Accepted: *26 July 2021*
Published: *19 August 2021*

Citation:
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

INTRODUCTION

Sea urchin is an ancient group of marine invertebrates that belongs to the class of Echinoidea under phylum of Echinodermata. To date, about 850 species of sea urchins have been found worldwide (Harris and Eddy, 2015). *Mesocentrotus nudus*, a member of the family Strongylocentrotidae, inhabits the intertidal and subtidal rocky sea bottoms along the coast of northwestern Pacific (Agatsuma, 2020). *M. nudus* is a well-known edible sea urchin for its good taste and redundant nutrition of vitamins and unsaturated fatty acids in the gonad (Mi et al., 2014). Furthermore, sea urchin extracts have extensive biological effects, such as anticancer, antioxidant, antileukemia, anti-fatigue, and anti-inflammatory effects (Liu et al., 2007).

The high edible and medicinal values of *M. nudus* have greatly stimulated the consumption demand, which has led to serious overfishing of wild sea urchins (Andrew et al., 2002; Li and Qi, 2008). In addition, the natural habitat of *M. nudus*, mostly adjacent to the densely populated areas (Adachi et al., 2020), has also been heavily disturbed by human activities including land-sourced pollution, aquaculture pollution, and sea reclamation (Lukyanova et al., 2017; Song and Duan, 2019). Under the influence of overfishing and natural habitat loss, the wild population resources of *M. nudus* are declining rapidly in recent years (Andrew et al., 2002). Especially in China, *M. nudus* has been considered as an endangered species. Although the breeding of *M. nudus* has been implemented to solve the shortage problem of natural resource in China since the 1980s (Ding et al., 2007), the parents of the seedlings were mainly from the wild populations. Therefore, understanding the genetic attributes of wild populations of *M. nudus* not only provide basic information for resources assessment and management and but also can lay a foundation for germplasm improvement.

Molecular markers have become the preferred tools for population genetic analysis. However, previously, quite limited genetic information on molecular markers (e.g., AFLP, mtDNA, and SSR) for *M. nudus* were available; and the level and pattern of population genetic diversity and structure are still poorly understood (Zhang et al., 2012). Recently, advances in next-generation sequencing (NGS) have greatly reduced the cost of nucleotide sequencing and made the genetic marker discovery more convenient. Genotyping-by-sequencing (GBS) as a genomic approach can explore the data covering the whole genome range with reduced cost and also provide more high-resolution genetic information for species lacking genomic information (Elshire et al., 2011). So far, GBS has been widely applied to assess genetic diversity on marine organisms, e.g., *Ostrea lurida* (Silliman, 2019), *Sillago japonica* (Yang et al., 2020), and *Haliotis discus* (Wells and Dale, 2018). However, no report on the population genetic analysis of *M. nudus* by using GBS was found up to now.

In the present study, GBS was used to genotype a total of 80 individuals consisting of five populations, covering the distribution range of *M. nudus* in China seas. The objectives were to detect and genotype single-nucleotide polymorphisms (SNPs) at a genome-wide level and to characterize the genetic diversity and population structure of *M. nudus*. The results will provide useful information for the utility of GBS on the genetic analysis of sea urchin species and contribute to the population conservation and genetic improvement of *M. nudus*.

MATERIALS AND METHODS

Sampling

In 2020, a total of 80 individuals were collected from five localities by SCUBA diving (16 individuals from each locality, **Figure 1**), covering the natural distribution range of *M. nudus* in China. Gonad tissue was sampled and fixed by ethanol on site. Genomic DNA of each sample was isolated from gonad tissue by using Plant Genomic DNA Kit (TIANGEN, Beijing, China) following the manual instruction. The purity and integrity of extracted DNA were determined by using a NanoDrop 1000 Spectrophotometer (NanoDrop, Wilmington, DE, United States) and electrophoresis on 1% agarose gel. Qualified genomic DNA was stored at −20°C.

Sequencing and Genotyping

All samples were sequenced by using the GBS (Qi et al., 2018) method, which was performed by OE Biotech Co., Ltd. (Shanghai, China). Briefly, extracted DNA from each individual was digested with restriction enzymes *Pst*I-HF and *Msp*I, and the barcoded adapters and common adapter were respectively ligated on the *Pst*I cut site and the *Msp*I cut site of all samples by T4 DNA ligase. Then, the fragments of 300–700 bp were retrieved using recovery system of improved magnetic bead and amplified by PCR using high-fidelity enzymes. Finally, amplification fragments were sequenced on an Illumina Nova, PE150 (Illumina, Inc., San Diego, CA, United States). The generated raw reads were filtered to remove the potential adaptor sequences by using STACKS software (Catchen et al., 2013). After quality control, there was no reference genome available for aligning the clean reads for SNP discovery. Therefore, ustacks (-M 3), cstacks (-n 3), sstacks, tsv2bam, and gstacks programs using STACKS pipeline were used for *de novo* SNP discovery. The initial SNPs were filtered according to the following parameters: (1) minimum sequence depth was above 5; (2) the allelic number was equal to 2; (3) minor allele frequency (MAF) was greater than 0.05, and (4) missing rate was lower than 10%. The quality control of SNP data was performed using vcftools (Danecek et al., 2011) and R software (R Core Team, 2018).

Genetic Diversity Analysis

Population genetic diversity within overall populations and within each of the five populations was assessed. The polymorphic information content (PIC) was calculated using R package snpReady (Granato et al., 2018). The observed heterozygosity (Ho), expected heterozygosity (He), and inbreeding coefficient (F_{is}) were calculated at each SNP locus, and also across all loci, using hierfstat package in R software (Goudet, 2005). Besides, multilocus heterozygosity (MLH) and standardized MLH (sMLH) were calculated for all individuals using inbreedR package (Stoffel et al., 2016). The effective population size (Ne) using the linkage disequilibrium (LD) method was estimated by using NeEstimator V2.01 (Do et al., 2014). Considering the small size of each population, the Ne was only analyzed for overall populations.

Population Structure Analysis

Initial analysis of population structure was carried out by principal component analysis (PCA) using R package SNPRelate (Zheng et al., 2012) and plotted using Scatterplot3d package (Ligges and Mächler, 2002). In addition, the levels of ancestry and admixture proportions of individuals were assessed by using R package LEA (Frichot and François, 2015). The number of ancestral populations (k) ranging between 1 and 10 were tested, and 10 replicate analyses were run for each value of k. The optimal k was chosen based on cross-entropy and cross-validation errors

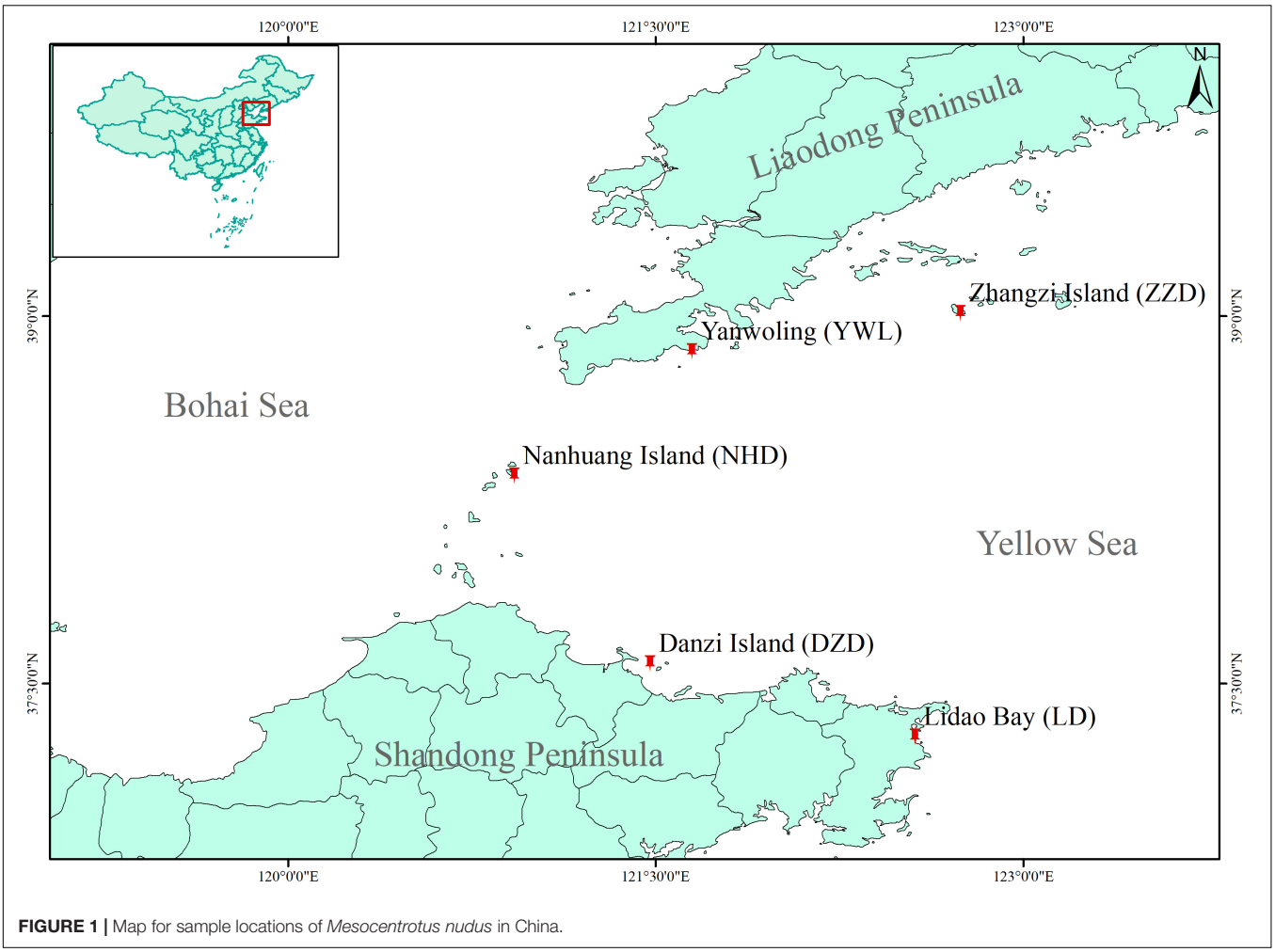

FIGURE 1 | Map for sample locations of *Mesocentrotus nudus* in China.

(Frichot et al., 2014). To better understand fine-scale patterns of genetic structure between populations, pairwise F_{st} values were calculated with the function pairwise.neifst and significance was tested using 1,000 bootstrap replicates with the function boot.ppfst in the R package hierfstat (Goudet, 2005).

Phylogenetic Analysis

The individual-level phylogenetic tree was constructed using SNPhylo (Lee et al., 2014) with an automated bash shell script snphylo.sh. For this analysis, an LD threshold (r^2 = 0.8) was used to reduce SNP redundancy, and 1,000 bootstrap analyses were fulfilled. The individual phylogenetic tree was visualized by using R package ggtree (Yu et al., 2017). Besides, the population-level phylogeny using the maximum likelihood approach was also inferred by using TreeMix (Pickrell and Pritchard, 2012). Considering that the missing data can affect the accuracy of TreeMix inference, the SNPs with missing values were removed. The population-level phylogeny was drawn by using plotting_funcs.R script[1].

[1]https://github.com/joepickrell/pophistory-tutorial/blob/master/example2/plotting_funcs.R

RESULTS

Sequencing and Genotyping

All samples were sequenced by using GBS at efficient sequencing depths of approximately 34.11-fold. In total, 76.76 Gb of raw data was produced, with an average of 0.96 Gb per individual. After filtering, a total of 72.39 Gb of clean data was retained with an average effective rate (clean reads compared with raw reads) of 94.30%. Statistics on clean reads further showed the quality value 20 (Q20) ≥ 96.33% and quality value 30 (Q30) ≥ 90.41%, and the guanine–cytosine (GC) contents ranged from 40.43 to 42.07%. The reference sequence constructed using STACKS pipeline contained 439,316 tags with an average length of 277 bp, the longest one being 1,036 bp and the shortest one being 71 bp. Initially, 992,728 putative SNPs were identified. After quality control based on successive filtering steps for sequence depth, allelic number, MAF, and missing rate, a total of 51,738 biallelic SNP markers located on 15,562 tags were retained.

Genetic Diversity

Several diversity indices are shown in **Table 1**. The PIC ranged from 0.213 (LD) to 0.220 (DZD), with an average of 0.217. The

TABLE 1 | Genetic diversity summary of *Mesocentrotus nudus* from five localities in China.

	PIC	Ho	He	F_{is}	MLH	sMLH	Ne
DZD	0.220	0.239	0.276	0.135	0.246	1.005	–
NHD	0.219	0.238	0.275	0.135	0.244	1.000	–
LD	0.213	0.230	0.266	0.135	0.244	1.006	–
ZZD	0.215	0.232	0.270	0.141	0.241	0.988	–
YWL	0.217	0.236	0.273	0.133	0.246	1.004	–
ALL	0.217	0.235	0.272	0.136	0.244	1.001	724.3

PIC, polymorphism information content; Ho, observed heterozygosity; He, expected heterozygosity; F_{is}, inbreeding coefficient; MLH, multilocus heterozygosity; sMLH, standardized MLH; Ne, effective population size.

observed heterozygosity (Ho) and expected heterozygosity (He) were in the ranges Ho = 0.230 to 0.239 and He = 0.266 to 0.276, with an average of Ho = 0.235 and He = 0.272, respectively. Regarding the inbreeding coefficient (F_{is}), the lowest value was detected in the YWL and the highest in ZZD. Average individual MLH ranged from 0.241 to 0.246. Average sMLH values slightly varied across populations. The estimated effective population size (Ne) was 724.3 for all populations.

Population Genetic Structure

The PCA revealed that the majority of samples were clustered together, and no geographical patterns were observed (**Figure 2**). Similar results can be achieved from the admixture analysis (**Figure 3**). With the use of LEA, the minimum value of the cross-validation errors was 0.710 when k = 2, and the values continuously increased with k from 3 to 10 (**Figure 3A**). All individuals were classified into two groups at the level of k = 2, but one group contained only two individuals from YWL (**Figure 3B**). To further investigate the population structure, the analysis at k = 3 was also performed, and similar results were observed (**Figure 3C**). The pairwise F_{st} values ranged from 0.0001 to 0.0039 with a mean value of 0.0019 (**Table 2**). The highest level of differentiation was observed between LD and YWL, whereas DZD and NHD differentiated the lowest.

Phylogenetic Tree

The individual-level phylogenetic tree was constructed with 33,178 biallelic SNPs (**Figure 4**), in which all individuals belong to the same branch except YWL8 and YWL10. The population-level phylogeny constructed with 5,460 biallelic SNPs showed that the five populations appeared to originate from the same group, and three migration events were modeled among different branches (**Figure 5**). Overall, the phylogenetic tree analysis coincided with the PCA and admixture analysis, which corroborated a high genetic connectivity with even biological mixing among all locations.

DISCUSSION

Reduced-representation sequencing could provide an opportunity for those species without prior genetic resources to understand their basic properties of the genome (Silliman, 2019; Yang et al., 2020). In this study, a large number of genome fragments were obtained for *M. nudus* by using GBS. By analyzing the DNA base composition, the percentage of GC

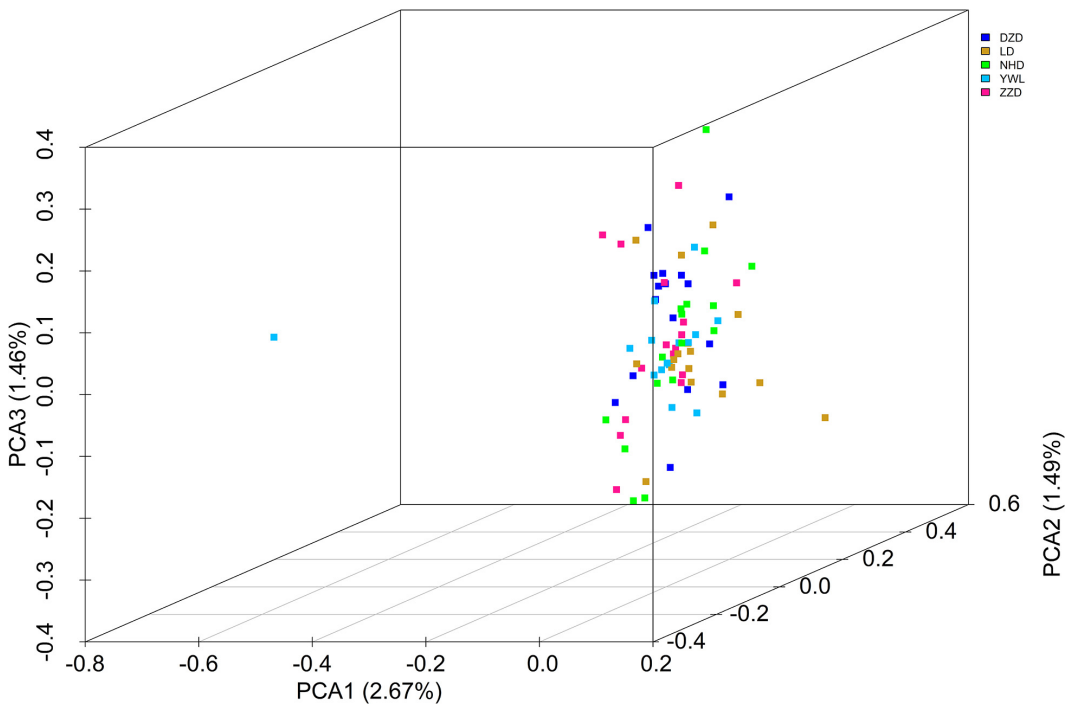

FIGURE 2 | Principal components analysis (PCA) 3D plot of the 80 individuals using genome-wide single-nucleotide polymorphisms (SNPs).

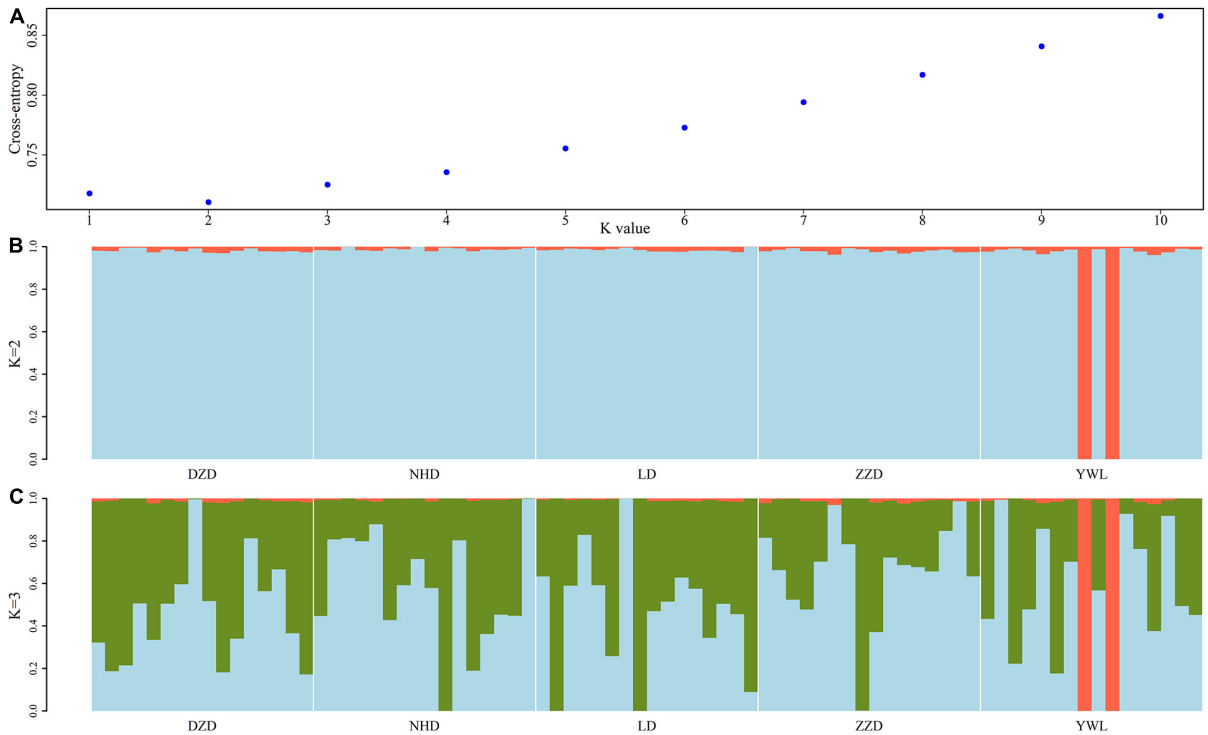

FIGURE 3 | Admixture analyses of 80 individuals based on the genome-wide single-nucleotide polymorphisms (SNPs). **(A)** Cross-validation plot for the number of ancestral populations (k). **(B)** Stacked bar plot for the k = 2. **(C)** Stacked bar plot for the k = 3. Each column represents an individual, and the length of colored segments denotes the proportion of an individual's genome inherited from one of k ancestral populations.

content was up to 41%, which was higher than that of the reported sea urchin, including *Strongylocentrotus purpuratus* (36.9%) (Sodergren et al., 2006), and *Lytechinus variegatus* (36.1%) (Davidson et al., 2020). Considering that genomic GC content is closely associated with genome size and significantly affects the genome functioning and species ecology (Šmarda et al., 2014), this result suggested that *M. nudus* may have a more complex genome than *S. purpuratus* and *L. variegatus*.

This study is the first report to identify genome-wide SNPs for *M. nudus* and to determine the genetic diversity and population structure of this species in China seas. Several indices concerning heterozygosity were firstly analyzed across five geographical populations. In general, the DZD population displayed the highest genetic diversity. This result may benefit from the good ecological status in the surrounding waters of Danzi island especially after removal of the marine raft culture and minimizing the amount of waste water discharged into the coastal waters since 2010 (Li et al., 2013). In contrast, the Lidao Bay as an important fishery and aquaculture water was seriously affected by anthropogenic disturbance (Jiang et al., 2013; Wu et al., 2016), which may result in the lowest PIC, Ho, and He for LD population. Besides, compared with previous studies on the population genetics of aquatic animals by using genome-wide SNPs, a moderate level of heterozygosity was observed for *M. nudus* in the current study. For example, the average Ho in *M. nudus* was lower than that in *Oncorhynchus keta* (Waples et al., 2017) and *Paracentrotus lividus* (Carreras et al., 2020), but higher than in several species such as *Oreochromis urolepis* (Nyinondi et al., 2020), *O. lurida* (Silliman, 2019), and *Isocladus armatus* (Wells and Dale, 2018). However, positive F_{is} values have been observed in *M. nudus*. F_{is} values could reflect the deviation from the Hardy–Weinberg equilibrium (HWE) genotype frequencies and indirectly reflect relative population heterozygosity. Generally, positive F_{is} values may be caused by mating among relatives, life history trait with free-spawned planktonic sperm, and selection acting on the genetic markers (Zouros and Foltz, 1984; David et al., 1997; Whitaker, 2004; Addison and Hart, 2005). In the current study, the estimated Ne value for *M. nudus* was significantly lower than that for other marine species such as *Chinook salmon* (Larson et al., 2014), *Salmo salar* (Johnston et al., 2014), and *Panulirus homarus* (Al-Breiki et al., 2018). According to the proposal by Frankham et al. (2014), this value is too low to

TABLE 2 | Pairwise F_{st} values among populations.

	DZD	NHD	LD	ZZD	YWL
DZD	–	–	–	–	–
NHD	0.0001	–	–	–	–
LD	0.0014	0.0018	–	–	–
ZZD	0.0003	0.0016	0.0012	–	–
YWL	0.0019	0.0032	0.0039	0.0031	–

No significant F_{st} values were found at $p < 0.05$.

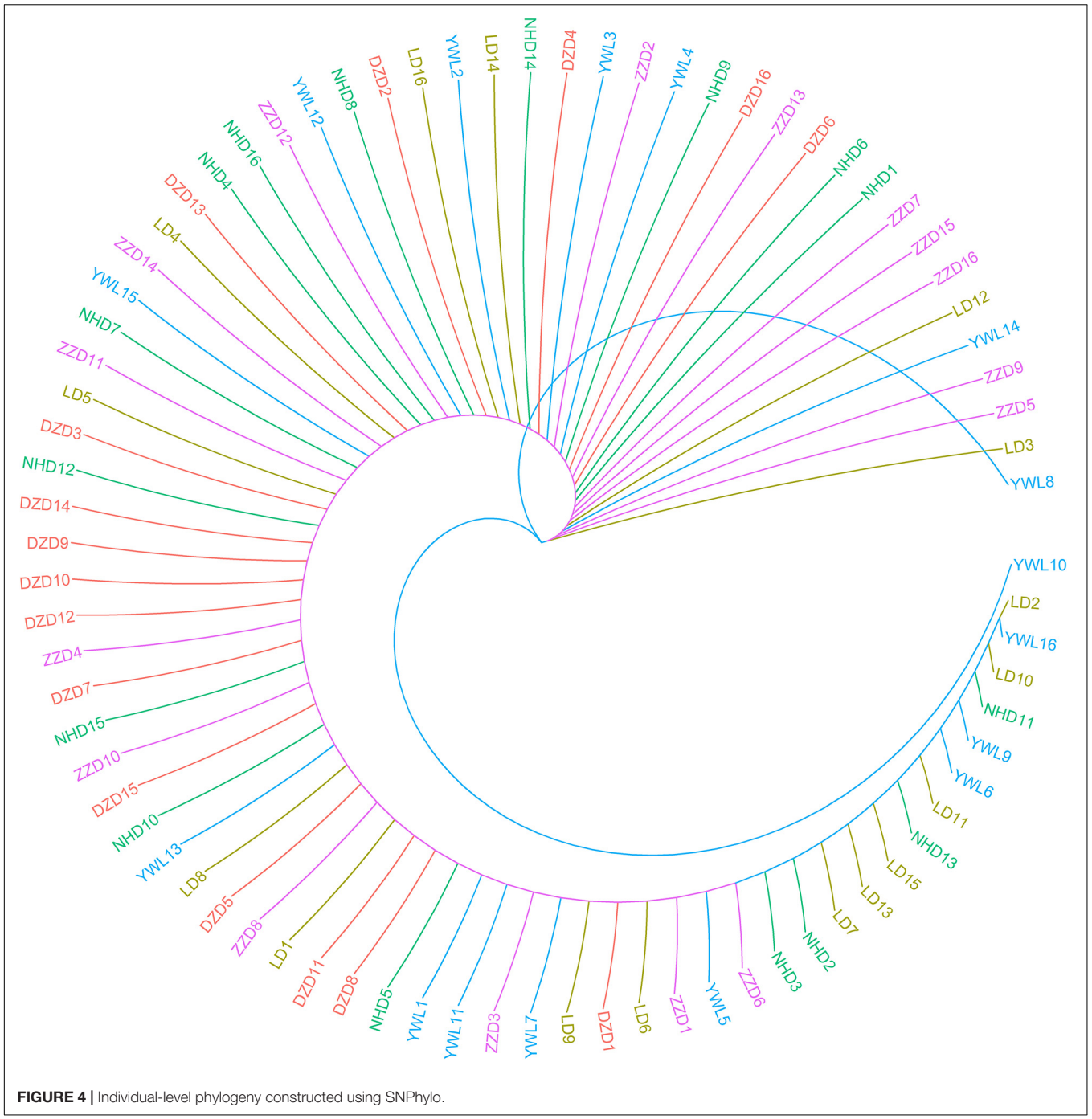

FIGURE 4 | Individual-level phylogeny constructed using SNPhylo.

retain evolutionary potential for *M. nudus* and thus may lead to inbreeding. Besides, the life history of *M. nudus* can also contribute to the positive F_{is} because the species like *M. nudus* with free-spawned planktonic sperm tend to have higher F_{is} than those species that copulate or have some form of direct sperm transfer to females or benthic egg masses (Addison and Hart, 2005). In addition, the population were sampled along coastal zone of the Yellow Sea and Bohai Sea, where they have been heavily disturbed by anthropogenic activities including land-source pollution, aquaculture pollution, and reclamation. The phenomena of eutrophication and hypoxia are happening with an increasing frequency in recent years, resulting in drastic changes in the physical and chemical parameters of seawater such as nutrient content, dissolved oxygen (DO), and pH especially in a small spatial scale. These drastic and persistent environmental disturbances can not only lead to the reduction of population size but also impose strong selective pressure on *M. nudus*, which is relatively weak in activity and thus may contribute to the high F_{is} value. Overall, the positive F_{is} and low Ne suggested that the genetic diversity of *M. nudus* should be

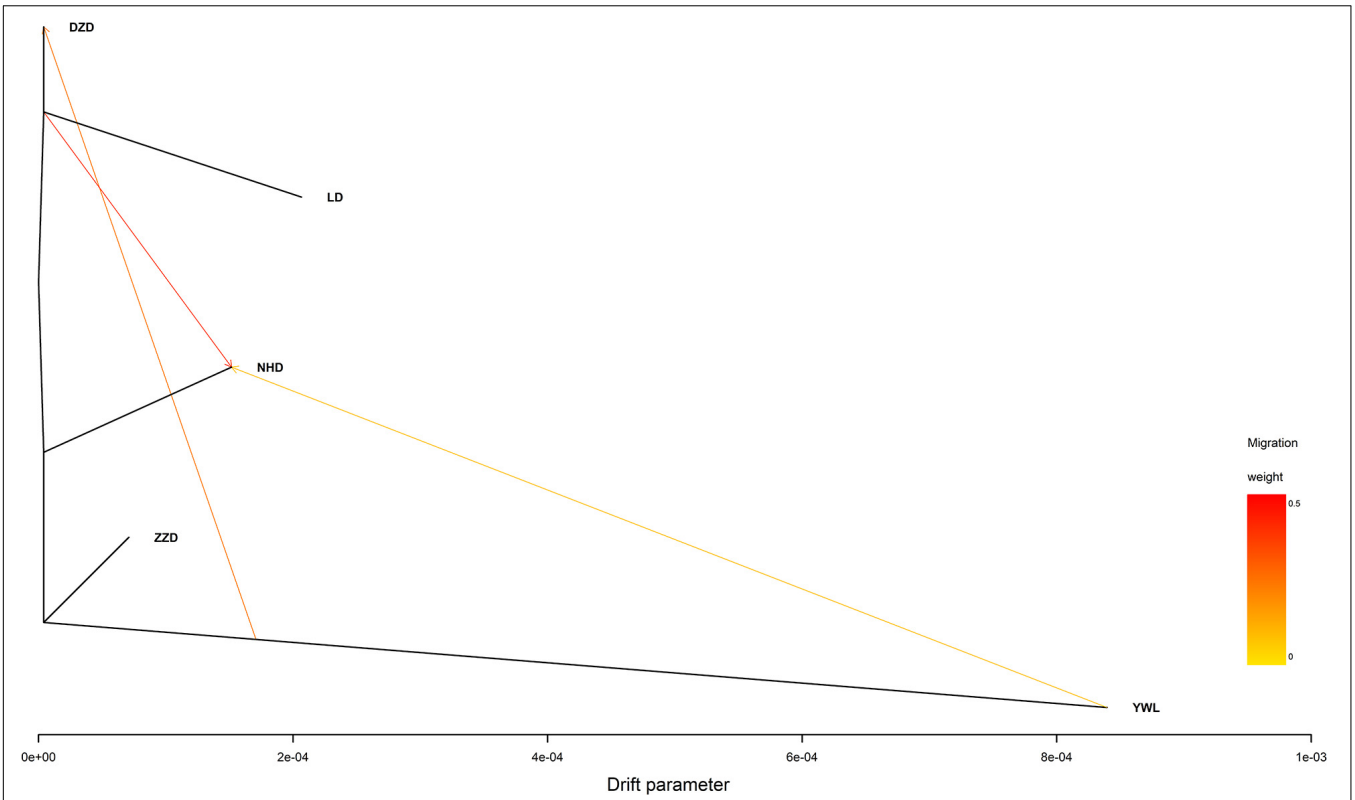

FIGURE 5 | Population-level phylogeny with migration edges constructed using TreeMix.

constantly monitored, and the protective measures should be carried out in China.

In order to assess the genetic differences among populations, pairwise F_{st} values were estimated. The results showed that all F_{st} values were far less than 0.05, which implies that there were low genetic differentiation and high genetic connectivity among five geographical populations. This result can be confirmed by PCA, admixture analysis, and phylogenetic analysis. This phenomenon may be mainly related to the lack of dispersal barriers in the Yellow Sea and Bohai Sea, and also the fact that *M. nudus* has a pelagic phase lasting for 1 month in the water column (Dautov et al., 2020). However, it is worth noting that two individuals from YWL deviated significantly from the main cluster and have little genetic association with other individuals. This phenomenon can also be found in similar studies. For example, individual ancestry analysis for five populations of Pacific cod showed that two individuals from coastal area had no similar ancestries as coastal fish (Drinan et al., 2018). Interestingly, these individuals displayed the same ancestry as the fish from the Salish Sea. Similarly, some seahorses from New Zealand were descendants of Australian ancestors (Ashe and Wilson, 2019). Therefore, in the current study, two differentiated individuals from YWL may imply that there is a genetically differentiated population in the adjacent sea. Especially, it is worth exploring whether the population structure of *M. nudus* from the Japan Sea will be different from that of the Yellow Sea and Bohai Sea, considering the two sea areas are separated by the Korean Peninsula, which might form a dispersal barrier. Besides, the small sampling size in the current study may be a potential factor leading to the cryptic genetic structure of *M. nudus* that may not be fully revealed. Therefore, the investigation with larger spatial scale and sample size of *M. nudus* needs to be implemented to fully assess the population structure and genetic diversity of *M. nudus* in the future.

CONCLUSION

In the present study, a total of 51,738 high-quality SNPs were detected, indicating that the GBS technology is a powerful tool for the genome-wide SNP discovery of *M. nudus*. Diversity index analysis showed that all populations have a similar pattern with positive F_{is} and low Ne, suggesting that it is necessary to carry out the conservation of *M. nudus* in China. Low genetic differentiation and high genetic connectivity among five geographical populations were detected by pairwise F_{st}, PCA, admixture analysis, and phylogenetic analysis. Besides, two individuals originating from an isolated ancestor were identified, which may be crucial for the conservation of genetic diversity and germplasm improvement. Thus, the investigation with larger spatial scale and sample size needs to be implemented to assess the population structure and genetic diversity of *M. nudus* 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 below: GenBank (Accession number: PRJNA738544).

ETHICS STATEMENT

Each of the procedures that were used to handle and treat the *Mesocentrotus nudus* during this study was approved by the Key Laboratory of Coastal Biology and Biological Resources Utilization, Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, prior to the initiation of the study and the Animal Management Regulations, revised on March 1, 2017, No. 676.

AUTHOR CONTRIBUTIONS

QW conducted the experiment and data processing. BL conceived and supervised the project. QW and YL collected the experimental animals. LY and LC contributed to preparing the genomic DNA for SNP genotyping. QW and BL wrote the manuscript. All authors have read and approved the manuscript.

FUNDING

This work was supported by Strategic Priority Research Program of the Chinese Academy of Sciences (XDA23050304), International Science Partnership Program of the Chinese Academy of Sciences (133137KYSB20200002), and Yantai Science and Technology Planning Project (2020YT06000390).

REFERENCES

Adachi, K., Suzuki, T., Okumura, S. I., Funayama, S., and Moriyama, S. (2020). Influence of the 2011 Tohoku tsunami on the genetic structure of wild sea urchin (*Mesocentrotus nudus*) populations in Sanriku, Japan. *Mar. Ecol.* 41:e12584. doi: 10.1111/maec.12584

Addison, J., and Hart, M. (2005). Spawning, copulation and inbreeding coefficients in marine invertebrates. *Biol. Lett.* 1, 450–453. doi: 10.1098/rsbl.2005.0353

Agatsuma, Y. (2020). *Mesocentrotus nudus*. *Dev. Aquac. Fish. Sci.* 43, 627–641.

Al-Breiki, R. D., Kjeldsen, S. R., Afzal, H., Al Hinai, M. S., Zenger, K. R., Jerry, D. R., et al. (2018). Genome-wide SNP analyses reveal high gene flow and signatures of local adaptation among the scalloped spiny lobster (*Panulirus homarus*) along the Omani coastline. *BMC Genomics* 19:690. doi: 10.1186/s12864-018-5044-8

Andrew, N., Agatsuma, Y., Ballesteros, E., Bazhin, A., Creaser, E., Barnes, D., et al. (2002). Status and management of world sea urchin fisheries. *Oceanogr. Mar. Biol.* 40, 343–425.

Ashe, J. L., and Wilson, A. B. (2019). Navigating the southern seas with small fins: genetic connectivity of seahorses (*Hippocampus abdominalis*) across the Tasman Sea. *J. Biogeogr.* 47, 207–219. doi: 10.1111/jbi.13733

Carreras, C., García-Cisneros, A., Wangensteen, O. S., Ordóñez, V., Palacín, C., Pascual, M., et al. (2020). East is east and west is west: population genomics and hierarchical analyses reveal genetic structure and adaptation footprints in the keystone species *Paracentrotus lividus* (Echinoidea). *Divers. Distrib.* 26, 382–398. doi: 10.1111/ddi.13016

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354

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

Dautov, S., Dautova, T., and Kashenko, S. (2020). Towards a scientific-based farming of sea urchins: first steps in the cultivation of *Diadema setosum*, *Diadema savignyi* and *Mesocentrotus nudus*. *APN Sci. Bull.* 10, 109–118. doi: 10.30852/sb.2020.1284

David, P., Perdieu, M. A., Pernot, A. F. O., and Jarne, P. (1997). Fine-grained spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. *Evolution* 51, 1318–1322. doi: 10.2307/2411061

Davidson, P. L., Guo, H., Wang, L., Berrio, A., Zhang, H., Chang, Y., et al. (2020). Chromosomal-level genome assembly of the sea urchin *Lytechinus variegatus* substantially improves functional genomic analyses. *Genome Biol. Evol.* 12, 1080–1086. doi: 10.1093/gbe/evaa101

Ding, J., Chang, Y., Wang, C., and Cao, X. (2007). Evaluation of the growth and heterosis of hybrids among three commercially important sea urchins in China: *Strongylocentrotus nudus*, *S. intermedius* and *Anthocidaris crassispina*. *Aquaculture* 272, 273–280. doi: 10.1016/j.aquaculture.2007.07.231

Do, C., Waples, R. S., Peel, D., Macbeth, G., Tillett, B. J., and Ovenden, J. R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol. Ecol. Resour.* 14, 209–214. doi: 10.1111/1755-0998.12157

Drinan, D. P., Gruenthal, K. M., Canino, M. F., Lowry, D., Fisher, M. C., and Hauser, L. (2018). Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*). *Evol. Appl.* 11, 1448–1464. doi: 10.1111/eva.12639

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One* 6:e19379. doi: 10.1371/journal.pone.0019379

Frankham, R., Bradshaw, C. J. A., and Brook, B. W. (2014). Genetics in conservation management: revised recommendations for the 50/500 rules, red list criteria and population viability analyses. *Biol. Conserv.* 170, 56–63. doi: 10.1016/j.biocon.2013.12.036

Frichot, E., and François, O. (2015). LEA: an R package for landscape and ecological association studies. *Methods Ecol. Evol.* 6, 925–929. doi: 10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., and François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics* 196, 973–983. doi: 10.1534/genetics.113.160572

Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol. Ecol. Notes* 5, 184–186. doi: 10.1111/j.1471-8286.2004.00828.x

Granato, I. S. C., Galli, G., de Oliveira Couto, E. G., e Souza, M. B., Mendonça, L. F., and Fritsche-Neto, R. (2018). Snpready: a tool to assist breeders in genomic analysis. *Mol. Breed.* 38:102. doi: 10.1007/s11032-018-0844-8

Harris, L. G., and Eddy, S. D. (2015). "Sea urchin ecology and biology," in *Echinoderm Aquaculture*, eds N. P. Brown and S. D. Eddy (Hoboken, NJ: John Wiley & Sons), 3–24.

Jiang, Z., Fang, J., Mao, Y., Han, T., and Wang, G. (2013). Influence of seaweed aquaculture on marine inorganic carbon dynamics and sea-air CO2 flux. *J. World Aquac. Soc.* 44, 133–140. doi: 10.1111/jwas.12000

Johnston, S. E., Orell, P., Pritchard, V. L., Kent, M. P., Lien, S., Niemelä, E., et al. (2014). Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Mole. Ecol.* 23, 3452–3468. doi: 10.1111/mec.12832

Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., and Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evol. Appl.* 7, 355–369. doi: 10.1111/eva.12128

Lee, T.-H., Guo, H., Wang, X., Kim, C., and Paterson, A. H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15:162. doi: 10.1186/1471-2164-15-162

Li, B., Wang, Q., and Li, B. (2013). Assessing the benthic ecological status in the stressed coastal waters of Yantai, Yellow Sea, using AMBI and M-AMBI. *Mar. Pollut. Bull.* 75, 53–61. doi: 10.1016/j.marpolbul.2013.08.007

Li, J., and Qi, L. (2008). A set of microsatellite markers for use in the endangered sea urchin *Strongylocentrotus nudus* developed from *S. purpuratus* ESTs. *Conserv. Genet.* 9, 743–745. doi: 10.1007/s10592-007-9382-3

Ligges, U., and Mächler, M. (2002). Scatterplot3d-an r package for visualizing multivariate data. *J. Statist. Softw.* 8, 1–20.

Liu, C., Lin, Q., Yi, G., Liang, Y., Xing, Y., and Tao, X. (2007). Characterization and antitumor activity of a polysaccharide from *Strongylocentrotus nudus* eggs. *Carbohydr. Polym.* 67, 313–318. doi: 10.1016/j.carbpol.2006.05.024

Lukyanova, O. N., Zhuravel, E. V., Chulchekov, D. N., and Mazur, A. A. (2017). Sea urchin embryogenesis as bioindicators of marine pollution in impact areas of the Sea of Japan/East Sea and the Sea of Okhotsk. *Arch. Environ. Contam. Toxicol.* 73, 322–333. doi: 10.1007/s00244-017-0388-7

Mi, X., Wei, Z., Zhou, Z., and Liu, X. (2014). Identification and profiling of sex-biased microRNAs from sea urchin *Strongylocentrotus nudus* gonad by solexa deep sequencing. *Comp. Biochem. Physiol. Part D Genom. Proteomics* 10, 1–8. doi: 10.1016/j.cbd.2014.01.001

Nyinondi, C. S., Mtolera, M. S., Mmochi, A. J., Lopes Pinto, F. A., Houston, R. D., de Koning, D. J., et al. (2020). Assessing the genetic diversity of farmed and wild Rufiji tilapia (*Oreochromis urolepis urolepis*) populations using ddRAD sequencing. *Ecol. Evol.* 10, 10044–10056. doi: 10.1002/ece3.6664

Pickrell, J., and Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8:e1002967. doi: 10.1038/npre.2012.6956.1

Qi, P., Gimode, D., Saha, D., Schröder, S., Chakraborty, D., Wang, X., et al. (2018). UGbS-Flex, a novel bioinformatics pipeline for imputation-free SNP discovery in polyploids without a reference genome: finger millet as a case study. *BMC Plant Biol.* 18:117. doi: 10.1186/s12870-018-1316-3

R Core Team (2018). *R: A Language and Environment for Statistical Computing. Computing*. Vienna: R Core Team.

Silliman, K. (2019). Population structure, genetic connectivity, and adaptation in the *Olympia oyster* (*Ostrea lurida*) along the west coast of North America. *Evol. Appl.* 12, 923–939. doi: 10.1111/eva.12766

Šmarda, P., Bureš, P., Horová, L., Leitch, I. J., Mucina, L., Pacini, E., et al. (2014). Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4096–E4102. doi: 10.1073/pnas.1321152111

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., et al. (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* 314, 941–952. doi: 10.1126/science.1133609

Song, J., and Duan, L. (2019). "The bohai sea," in *World Seas: An Environmental Evaluation*, 2nd Edn, ed. C. Sheppard (Cambridge, MA: Academic Press), 377–394.

Stoffel, M. A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., et al. (2016). Inbreedr: an R package for the analysis of inbreeding based on genetic markers. *Methods Ecol. Evol.* 7, 1331–1339. doi: 10.1111/2041-210X.12588

Waples, R., Seeb, J., and Seeb, L. (2017). Congruent population structure across paralogous and nonparalogous loci in Salish sea chum salmon (*Oncorhynchus keta*). *Mol. Ecol.* 26, 4131–4144. doi: 10.1111/mec.14163

Wells, S. J., and Dale, J. (2018). Contrasting gene flow at different spatial scales revealed by genotyping-by-sequencing in *Isocladus armatus*, a massively colour polymorphic New Zealand marine isopod. *PeerJ* 6:e5462. doi: 10.7717/peerj.5462

Whitaker, K. (2004). Non-random mating and population genetic subdivision of two broadcasting corals at Ningaloo Reef. Western Australia. *Mar. Biol.* 144, 593–603. doi: 10.1007/s00227-003-1220-7

Wu, Z., Zhang, X., Lozano-Montes, H. M., and Loneragan, N. R. (2016). Trophic flows, kelp culture and fisheries in the marine ecosystem of an artificial reef zone in the Yellow Sea. *Estuar. Coast. Shelf Sci.* 182, 86–97. doi: 10.1016/j.ecss.2016.08.021

Yang, T., Gao, T., Meng, W., and Jiang, Y. (2020). Genome-wide population structure and genetic diversity of Japanese whiting (*Sillago japonica*) inferred from genotyping-by-sequencing (GBS): implications for fisheries management. *Fish. Res.* 225:105501. doi: 10.1016/j.fishres.2020.105501

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T. Y. (2017). Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628

Zhang, L., Dong, Y., Yang, A., Chen, Z., Wang, X., Guan, X., et al. (2012). Genetic diversity analysis of three geographic populations of sea urchin *Strongylocentrotus nudus* by AFLP. *Fish. Sci.* 31, 132–136. doi: 10.16378/j.cnki.1003-1111.2012.03.007

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606

Zouros, E., and Foltz, D. W. (1984). Possible explanations of heterozygote deficiency in bivalve mollusks. *Malacologia* 25, 583–591.

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 in Genetics

ORIGINAL RESEARCH
published: 08 September 2021
doi: 10.3389/fgene.2021.704778

GH Overexpression Alters Spermatic Cells MicroRNAome Profile in Transgenic Zebrafish

*William B. Domingues[1], Tony L. R. Silveira[2], Leandro S. Nunes[1], Eduardo B. Blodorn[1], Augusto Schneider[3], Carine D. Corcine[4], Antônio S. Varela Junior[4], Izani B. Acosta[4], Mateus T. Kütter[2], Gonzalo Greif[5], Carlos Robello[5], Danillo Pinhal[6], Luís F. Marins[2] and Vinicius F. Campos[1]**

[1] Laboratório de Genômica Estrutural, Programa de Pós-Graduação em Biotecnologia, Centro de Desenvolvimento Tecnológico, Universidade Federal de Pelotas, Pelotas, Brazil, [2] Laboratório de Biologia Molecular, Instituto de Ciências Biológicas, Universidade Federal do Rio Grande, Rio Grande, Brazil, [3] Faculdade de Nutrição, Universidade Federal de Pelotas, Pelotas, Brazil, [4] ReproPel, Programa de Pós-Graduação em Veterinária, Faculdade de Veterinária, Universidade Federal de Pelotas, Pelotas, Brazil, [5] Unidad de Biología Molecular, Institut Pasteur, Montevideo, Uruguay, [6] Laboratório Genômica e Evolução Molecular Departamento de Genética, Instituto de Biociências de Botucatu Universidade Estadual Paulista (UNESP), Botucatu, Brazil

OPEN ACCESS

Edited by:
Zexia Gao,
Huazhong Agricultural University, China

Reviewed by:
Tara G. McDaneld,
U.S. Meat Animal Research Center, Agricultural Research Service, United States Department of Agriculture, United States
Jie Mei,
Huazhong Agricultural University, China

***Correspondence:**
Vinicius F. Campos
fariascampos@gmail.com

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

Received: *03 May 2021*
Accepted: *23 August 2021*
Published: *08 September 2021*

Citation:
Domingues WB, Silveira TLR, Nunes LS, Blodorn EB, Schneider A, Corcine CD, Varela Junior AS, Acosta IB, Kütter MT, Greif G, Robello C, Pinhal D, Marins LF and Campos VF (2021) GH Overexpression Alters Spermatic Cells MicroRNAome Profile in Transgenic Zebrafish. Front. Genet. 12:704778. doi: 10.3389/fgene.2021.704778

Overexpression of growth hormone (GH) in *gh*-transgenic zebrafish of a highly studied lineage F0104 has earlier been reported to cause increased muscle growth. In addition to this, GH affects a broad range of cellular processes in transgenic fish, such as morphology, physiology, and behavior. Reports show changes such as decreased sperm quality and reduced reproductive performance in transgenic males. It is hypothesized that microRNAs are directly involved in the regulation of fertility potential during spermatogenesis. The primary aim of our study was to verify whether *gh* overexpression disturbs the sperm miRNA profile and influences the sperm quality in transgenic zebrafish. We report a significant increase in body weight of *gh*-transgenic males along with associated reduced sperm motility and other kinetic parameters in comparison to the non-transgenic group. MicroRNA transcriptome sequencing of *gh*-transgenic zebrafish sperms revealed expressions of 186 miRNAs, among which six miRNA were up-regulated (miR-146b, miR-200a-5p, miR-146a, miR-726, miR-184, and miR-738) and sixteen were down-regulated (miR-19d-3p, miR-126a-5p, miR-126b-5p, miR-22a-5p, miR-16c-5p, miR-20a-5p, miR-126b-3p, miR-107a-3p, miR-93, miR-2189, miR-202–5p, miR-221–3p, miR-125a, miR-125b-5p, miR-126a-3p, and miR-30c-5p) in comparison to non-transgenic zebrafish. Some of the dysregulated miRNAs were previously reported to be related to abnormalities in sperm quality and reduced reproduction ability in other species. In this study, an average of 134 differentially expressed miRNAs-targeted genes were predicted using the *in silico* approach. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis demonstrated that the genes of affected pathways were primarily related to spermatogenesis, sperm motility, and cell apoptosis. Our results suggested that excess GH caused a detrimental effect on sperm microRNAome, consequently reducing the sperm quality and reproductive potential of zebrafish males.

Keywords: non-coding RNAs, epigenetic, *Danio rerio*, miRNA-seq, sperm motility, transgenic fish

INTRODUCTION

Since the first proposal to use zebrafish (*Danio rerio*) in research (Creaser, 1934), the species have attained global popularity as an experimental model organism (Streisinger et al., 1981), and its use in science has grown and continues to grow rapidly (Kinth et al., 2013; Teame et al., 2019). Approximately 32 wild strains, such as AB, Tübingen long-fin, and Tüpfel long-Fin, are currently used (which excludes the local pet shop variants) in various studies (van den Bos et al., 2020; Silveira et al., 2021).[1] These reports show more than 115,000 genetic alterations done in zebrafish, of which more than 47,000 are transgenic insertions (Ruzicka et al., 2019). The first transgenic fish, the F0104 strain, was developed in Brazil in 2004 (Figueiredo et al., 2007a). Through transgenesis, the gene sequences can be manipulated to assign characteristics of interest. In addition, animal cells can be artificially labeled for easy visualization (Ingham, 2009).

The F0104 strain zebrafish model used in this study carried two transgenes: (1) the β-actin promoter from carp (*Cyprinus carpio*) that drives the expression of coding DNA sequence (CDS) of growth hormone gene (*gh*) from the marine silverside *Odonthestes argentinensis* and (2) green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* used as a transgenesis label controlled by the same promoter. In related studies, these transgenic zebrafish showed to have increased growth (Figueiredo et al., 2007b); better oxygen consumption (Rosa et al., 2008; Almeida et al., 2013a); Reactive oxygen species (ROS) production (Rosa et al., 2008); cognition (Studzinski et al., 2015); and regeneration capacity (Nornberg et al., 2016). However, on the other side, these strains have also shown to develop osmotic and energy imbalance (Almeida et al., 2013b); weakened immunity (Batista et al., 2014); deficits in antioxidant defense mechanism (da Rosa et al., 2011); less apparent sexual dimorphism (Figueiredo et al., 2007a); decrease in spermatic parameters and reproductive capacity of males (Figueiredo et al., 2013).

In the male reproductive system, growth hormone (GH) is said to interfere with testicular development and stimulate spermatogenesis by acting both directly and indirectly on testes through potentiation of gonadotropin, affecting the gonadal development by stimulating the expression of insulin-like growth factors (IGFs) (Gac et al., 1993; Berishvili et al., 2006; Miura et al., 2011). In fact, in both cases, GH may be considered as a co-factor for gonadotropin (Hull and Harvey, 2002) along with IGFs to specifically regulate the gonadal function (Gac et al., 1993). Overexpression of *gh* in other transgenic fish strains culminates in reproductive deficits, such as in tilapia (Rahman et al., 1998); Atlantic salmon (Moreau et al., 2011); coho salmon (Devlin et al., 2004); mud loach (Nam et al., 2002); and common carp (Cao et al., 2014). Some authors highlight that the factors that link somatic growth with reproductive deficits in *gh*-transgenic fish still need to be elucidated (Chen et al., 2018). A modern tool available for this task is the analysis of differential expression of microRNAs (miRNAs).

MicroRNAs are a class of endogenous non-coding RNAs (ncRNA), generally about 22 nucleotides in length regulating gene expressions canonically at the post-transcriptional level (Blödorn et al., 2021). To date, records in miRBase (Kozomara et al., 2019) and MiRGeneDB databases (Fromm et al., 2020) report 373 and 390 mature miRNA molecules are being expressed in zebrafish, respectively. Estimates suggest that between 30 and 60% of the total gene expressions can be regulated by miRNAs (Lewis et al., 2005; Xie et al., 2005; Friedman et al., 2008). MiRNAs are involved in virtually all biological processes in the cells of metazoans, including the spermatogenesis and formation of spermatozoa in zebrafish (Kotaja, 2014; Jia et al., 2015). Due to its wide distribution among species and role in regulating gene expressions, miRNAs have been used as biomarkers for various applications, including infertility diagnosis (Wang et al., 2015; Corral-vazquez et al., 2019; dos Santos da Silva et al., 2021).

Despite high maintenance of the F0104 strain being a major issue, the inherent reproductive deficit presented in the F0104 males provides an opportunity to understand the effects of *gh* expression levels on the reproduction potential using advanced approaches. Therefore, the F0104 strain has been used as an interesting model for investigations on spermatic quality, reproductive potency, and their determining factors in males. In addition, zebrafish is a translational model, and therefore the results obtained in this species can be extrapolated to other species, including humans (Vargas, 2018). Thus, the present study aimed to evaluate the sperm kinetic parameters as well as the sperm quality by sequencing, identifying, and quantifying miRNAs in sperm cells of *gh*-transgenic (*gh*+) and non-transgenic (NT) zebrafish, further elucidating the specific pathways influenced by transgenesis and lastly, to identify prospective candidate miRNAs as epigenetic biomarkers for the reproductive deficit of zebrafish males.

MATERIALS AND METHODS

Zebrafish

The F0104 zebrafish strain used in the study were *gh*-transgenic males. The NT fish were siblings of the *gh* + zebrafish (having no genetic construct incorporated into its genome). Both transgenic and NT zebrafish were selected in their larval stage, and 20 fish from each type were distributed into four different aquaria (2 for *gh* + and 2 for NT) kept in a closed water recirculation system. Each aquarium was maintained in 15 L water with oxygen levels near saturation (>6 mg/L), pH close to 7.2, photoperiod 14:10 h light: dark cycle, and temperature of 28°C. Ammonia (NH_3) and nitrite (NO_2^-) concentrations were measured three times a week using commercial kits (LabconTest Toxic Ammonia Freshwater; LabconTest Nitrite, Alcon®, Brazil) and corrected with partial water change when necessary. The levels of ammonia and nitrites showed to maintain below 0.011 and 0.25 mg/L, respectively. The fish were fed *ad libitum* twice a day with commercial fish food (ColorBits, Tetra®, Germany).

[1] https://www.zfin.org

Testes Excision

When the fish reached 15 months of age, which is normally considered adult age, presenting appropriate sperm production with normal sperm concentration and motility values (Johnson et al., 2018), 10 males from each experimental group were randomly captured by the same experienced operator (5 fish from each aquarium) and euthanized by hypothermic shock (in water with 4°C) for the removal of testes. One collected testis from each fish was placed in 100 μL of Beltsville Thawing Solution (BTS) (Pursel and Johnson, 1975) for semen analyses and the other preserved in TRIzol Reagent (Invitrogen, United States) for miRNA sequencing. This study was conducted in compliance with institutional, national, and international guidelines for using animals, and all the protocols used were performed by the guidelines and approved by the Ethics Committee of the Federal University of Rio Grande (FURG), Brazil, under the code 23116.008403/2018–32.

Sperm Kinetics Parameters

Sperm cells were assessed by Computer Assisted Sperm Analysis (CASA) system (SpermVision®, Minitube, Germany) coupled to an optical microscope (Axio Scope A1®, Zeiss, Germany) and observed at 200 X magnification according to the method described by Acosta et al. (2016). To activate sperm cells, activation solution consisting of 119 mM sodium bicarbonate solution ($NaHCO_3$) was prepared, added to semen in a 1:5 ratio, and plated on a slide. Variables assessed by CASA were: Total motility (TMO), Progressive motility (PMO), Velocity average path (VAP), Velocity curved line (VCL), Velocity straight line (VSL), and Linearity (LIN). For each sample, at least 500 cells were observed in 10 different fields.

The duration of sperm motility (in seconds) was evaluated using semen sample aliquots in 1:5 ratio using activation solution (prepared as earlier) and loaded on slides covered with a coverslip, a timer set, and time noted until sperm cells movement ceased.

RNA Isolation From Sperm Cells

Total RNA was isolated from sperm cells using a method that combined TRIzol reagent (Invitrogen, United States) and RNeasy mini spin column (RNeasy Mini Kit, Qiagen Inc., Valencia, CA, United States), according to the manufacturer's instructions with modifications. Briefly, testes samples were placed in 1 mL lysis solution containing 0.5% Triton and 0.1% SDS at room temperature for 30 min for removal of the contaminating somatic cells. The digested tissue solution was centrifuged at 50 × g for 5 min, and tissue debris in the pellet was discarded. To the supernatant, about 500 μL TRIzol reagent was added and thoroughly mixed. A known volume of chloroform was added to the homogenate, mixed vigorously, and incubated for 10 min at room temperature. The mixture was then centrifuged at 12,000 × g at 4°C for 15 min. After centrifugation, the upper aqueous phase was collected and mixed with an equal volume of 100% ethanol. The entire mixture was transferred to the RNeasy mini spin column. The RNA bound to the membrane of the spin column when centrifuged was subsequently removed using buffer RWT and buffer RPE and eluted in 30 μL RNase-free water. Concentration and purity of the isolated sperm RNA were measured using Agilent 4,200 TapeStation system and the Agilent RNA ScreenTape assay (Agilent Technologies, Santa Clara, United States) (**Supplementary Figure 1**). All RNA samples were stored at −80°C until used.

MicroRNA Sequencing

Small RNA libraries of sperm cells were generated using Next Small RNA Library Prep Set (New England Biolabs, Inc., United States) for the Illumina platform. To prepare each library set, three sperm RNA samples in equal concentrations were pooled together. A total of eighteen samples were used for the small RNA library preparation process, which culminated in 3 libraries for the *gh* + group and 3 libraries for the NT group. The manufacturer's instructions were followed throughout for the PCR amplification (15 cycles), taking 1,000 ng template as starting input. Following PCR amplification, the libraries were run in a 6% polyacrylamide gel, and the ∼140 bp bands were excised from the gels using a razor blade and quantified by BioAnalyzer 2,100 (Agilent Technologies, United States).

The concentrations of the six RNA libraries (3 from *gh* + and 3 from NT groups) were determined using the Invitrogen Qubit™ dsDNA High Sensitivity Assay Kit. The libraries were pooled at equimolar concentrations before MiSeq (Illumina, United States) sequencing in 50-bp single-end reads configuration using v3 sequencing chemistry.

MicroRNA Detection and Differential Expression Analysis

Raw sequence data were subjected to a cleaning process to remove 5′ and 3′ primer contaminants, N adaptors, polyA adaptors, sequences without index sequence tag, and adaptors shorter than 17 nucleotides. The alignment of sequences from these libraries and identification of miRNAs was performed using the webserver RNA toolbox,[2] according to methods used by Schneider et al. (2018). Differential expression of miRNAs was analyzed using the package edgeR (version 3.34) for both the *gh* + and NT experimental groups. Those miRNAs with False Discovery Rate (FDR) > 0.05 and Fold Changes (FC) in expression > 2.0 were considered to be upregulated, while those with FDR < 0.05 and FC < 0.50 were considered to be downregulated.

Target Prediction and Gene Ontology Enrichment Analysis

The tool mirPath version 3.0 (Vlachos et al., 2015) was used to identify genes regulated by differentially expressed miRNAs, using the database microT-CDS v.5.0 (Paraskevopoulou et al., 2013). The gene ontology (GO) terms of the most enriched miRNA targets were annotated using the same mirPath tool. The gene targets of miRNAs were organized based on their functional role in different biological processes. The KEGG molecular pathways regulated by the differentially expressed

[2]https://bioinfo5.ugr.es/srnatoolbox

miRNAs (DEmiRNAs) with a p-value < 0.05 were considered as significantly enriched in our analysis.

Statistical Analysis

The statistical analyses of sperm kinetics parameters were performed using Statistix version 10.0. Compiled datasets were evaluated by Student's t-test, and the differences considered significant when $p < 0.05$. Data normality and homogeneity of variances were previously verified.

RESULTS

Effects on Body Weight and Spermatozoa Kinetic Parameters

The biometric measure of weight (**Figure 1B**) demonstrated that transgenesis (**Figure 1A**) led to a significant increase in body weight of *gh* + individuals when compared to NT individuals (544.6 ± 13.59 g and 466.6 ± 19.13 g, respectively).

The results obtained through the CASA to evaluate sperm motility demonstrated that *gh* + individuals had low total motility (**Figure 1C**) and lesser duration of spermatozoa motility (**Figure 1D**) when compared to the NT individuals ($P < 0.0001$). In addition, a significant decrease in other kinetic parameters was observed, such as reduced progressive motility and lower distances covered, both of which relate to the fertilizing potential of sperm in these transgenic individuals (**Table 1**).

Differential Expression of miRNAs in *gh*-Transgenic Strain

The microRNAome sequencing generated 21,981,994 raw reads from the six miRNA libraries altogether. After clean-up, a total of 186 known mature miRNAs were identified. Expression analysis of these miRNAs revealed 22 differentially expressed miRNAs (DEmiRNAs) in the *gh* + sperm cells (**Table 2**), including 16 down-regulated and 6 upregulated miRNAs in comparison to NT cells (**Supplementary Table 1**). Interestingly, for the miR-126 paralogs, both arms were down-regulated, suggesting that their 5p and 3p mature transcripts are functionally associated, controlling distinct genes that take part in the same gene regulatory network.

Molecular Pathways Regulated by miRNAs

An average of 134 putative target genes for the 22 DEmiRNAs was predicted using microT-CDS database (**Table 2**). After enrichment analysis, nine KEGG pathways were identified to be significantly related to DEmiRNAs gene targets (**Table 3**). Accordingly, down-regulated miRNAs were shown to be enriched for the FoxO signaling pathway, which plays a role in protein processing in the endoplasmic reticulum, sulfur relay system, apoptosis, and associated with the p53 signaling pathway. On the other hand, the upregulated miRNAs were shown to be enriched for pathways involved in mucin-type O-Glycan biosynthesis, glycosaminoglycan biosynthesis-heparan sulfate/heparin, N-Glycan biosynthesis, Toll-like receptor signaling, and glycerophospholipid metabolism. All these observations together indicate that the GH overexpression affects gene expression of key sperm motility pathways by altering the levels of regulatory miRNAs.

DISCUSSION

To the best of our knowledge, the present study is the first to demonstrate the effects of GH overexpression on the microRNAome profile of sperm cells in *gh* + transgenic zebrafish. We demonstrated that 16 miRNAs were down-regulated, and 6 miRNAs were upregulated in *gh* + transgenic fish when compared to the NT fish. Some of these candidates' differentially expressed miRNAs have been previously reported to play a potential role in influencing spermatic quality and reproductive success. Furthermore, a high level of sequence conservation in the aforementioned miRNAs of the metazoan genome helps up to directly compare the current findings in zebrafish (*Danio rerio*) with known findings of other species.

Among the down-regulated miRNAs identified, miR-20a-5p was earlier reported to express at lower levels in bovine sperm (Capra et al., 2017) and human blood plasma along with low total and progressive sperm motility (Cito et al., 2020). Likewise, the miR-202-5p is a highly expressed factor in mice testis during the early stages of development that re-surge during its final stages of development (Wainwright et al., 2013), posing to be strongly associated with the development of male gonad. Particularly in zebrafish, miR-202-5p shows high levels of expression in germ cells in the different stages of spermatogenesis, including the formation of spermatogonia and maturation of spermatozoa, thereby reinforcing the existence of a relationship between miR-202-5p and sperm maturation in this species (Jia et al., 2015).

In the present study, we observed the down-regulation of both miR-125a and miR-125b-5p expressions in the sperm cells of *gh* + zebrafish. We also reported in our earlier studies that these members of a miR-125 family were down-regulated when tested in nanotransfected sperm cells of *Bos taurus* shown by the negative changes in sperm kinetics and other parameters, such as its membrane integrity, acrosome reaction, and mitochondrial membrane potential (Domingues et al., 2020). A knockout study of miR-125b5p in mice showed it to cause male infertility (Li et al., 2019), indicating that the negative regulation of this miRNA in *gh* + zebrafish found here may also have deleterious effects on the fertilizing potential of the male F0104 zebrafish strain.

Further, the upregulated miR-146b observed in our study correlated with the previously observed abundant levels in bull testicular tissue (Gao et al., 2020), where its increased expression led to the inhibition of proliferation of germline stem cells and simultaneously promoted their apoptosis. Similarly, a deletion in the chromosomal locus for the miR-200a-5p precursor gene showed to increase the kinetics parameters of spermatozoa and significantly improved the fertilization rate in zebrafish in a different study (Xiong et al., 2018). This effect was caused due to the ectopic expression of miR-200a-5p by reducing expressions of its target genes: *wt1a*, *srd5a2b,* and *amh,* which are known to play an important role in spermatozoa motility.

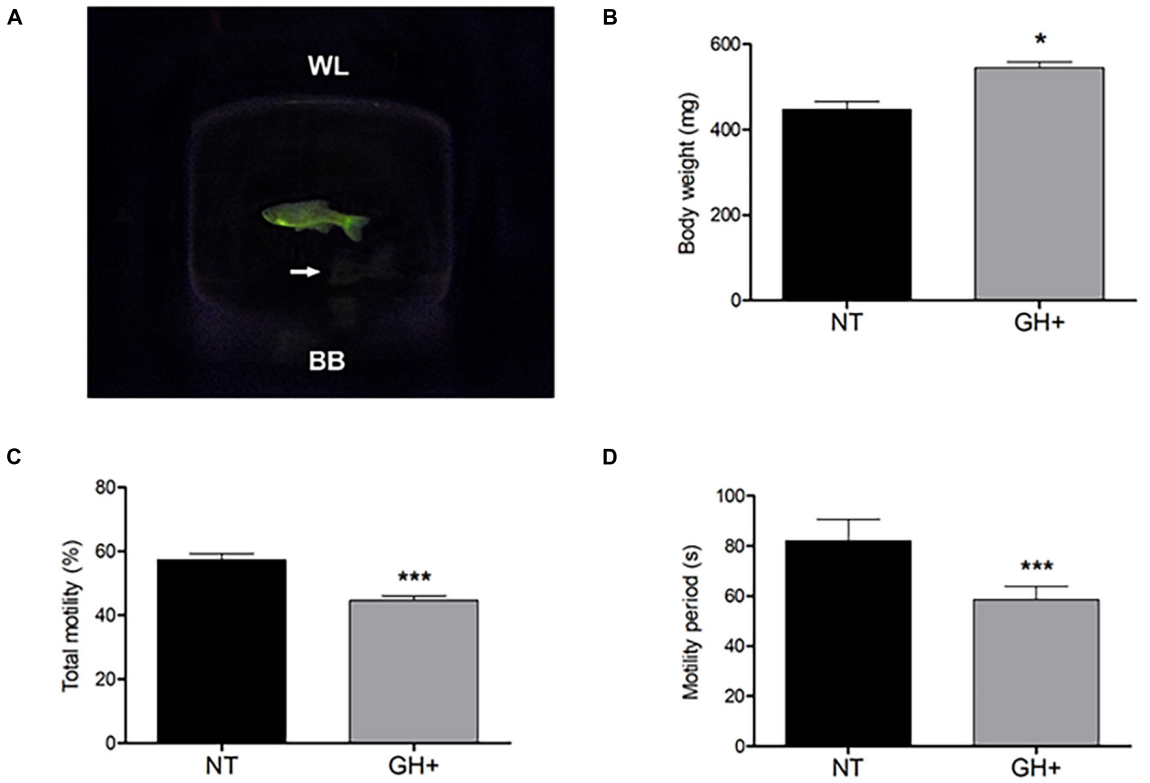

FIGURE 1 | Indications of transgenic phenotype in a photographic image of male zebrafish, changes in body weight, and sperm kinetic parameters of *gh*-transgenic F0104 strain zebrafish (*Danio rerio*) and non-transgenic strain. **(A)** *gh*-transgenic zebrafish male from F0104 strain showing the ubiquitous and constitutive distribution of GFP expression; non-transgenic zebrafish shown in white arrows; **(B)** bodyweight characteristics; **(C)** total motile sperm; **(D)** sperm motility period. All *p*-values < 0.05 are summarized with one asterisk and < 0.0001 with three asterisks. Results are statistically expressed by mean ± standard error of the mean (SEM) ($n = 10$ for each experimental group). NT, non-transgenic; GH+, *gh*-transgenic; WL, water line; BB, bottom of beaker.

The miR-184 was reported to be upregulated in the low motility fraction of bull sperms (Capra et al., 2017). In humans, the miR-184 plays an important role in male fertility by regulating genes related to spermatogenesis (Rahbar et al., 2020). Overall, the upregulations of these miRNAs show to have a negative effect on male fertility, suggesting that this may be a key reason for decreased fertility in *gh* + zebrafish.

The remaining differentially expressed miRNAs identified in our studies, such as the miR-22a-5p, miR-126, miR-22a, and miR-221 have also been earlier reported in semen samples of zebrafish and other species (Smorag et al., 2012; Jia et al., 2015; Tao et al., 2018; Alves et al., 2021). However, the functional role of these miRNAs and target mRNAs has not yet been fully elucidated. Further, characterization of these candidate miRNAs influencing the zebrafish spermatozoa characteristics and reproductive success is opportunely needed.

The *in silico* analysis of molecular pathways affected by differentially expressed miRNAs confirmed that the *gh* overexpression may be influencing the gene expressions related to pathways associated with spermatic motility. The FoxO signaling pathway, which includes transcription factors regulating the gene expressions related to apoptosis, cell cycle control, glucose metabolism, and resistance to oxidative stress identified in our study, was one of the main target sites for the down-regulated miRNAs (Wang et al., 2014). Therefore, a reduction in the miRNAs targeting *foxo3a* suggests a positive regulation of this transcription factor, consequently activating the FoxO signaling pathway. In addition, several other genes downstream from the coding region for *foxo3a* transcription factor have also shown to be targeted by the down-regulated miRNAs, further implying their importance in this pathway (by themselves or in association with PI3K/Akt signaling pathway) in spermatic cells of *gh* + zebrafish (Griffeth et al., 2013; Capra et al., 2017; Zhang et al., 2017b; Catalano-Iniesta et al., 2019).

In this study we observed an upregulation of miRNAs that affects genes related to Toll-like receptor signaling pathway in spermatic cells of gh-transgenic zebrafish. As discussed by Navarro-Costa et al. (2020), the functional significance of these immunity-related genes (such as *tlr7*, *tlr8, and traf6*) on the sperm motility process is not fully elucidated. Nevertheless, other study demonstrated that stimulated Toll-like receptor signaling reduce sperm motility and suppress fertilization in human (Fujita et al., 2011); Furthermore, a number of TLRs in testes of yellow catfish were significantly reduced by chloroquine treatment, resulting in an improved sperm motility and fertilization rate (Zhang et al., 2017a). Interestingly, in the present study, the affected main

TABLE 1 | Evaluation of sperm sample parameters in F0104 *gh*-transgenic and non-transgenic zebrafish (*Danio rerio*) by Computer Assisted Sperm Analysis (CASA) system: Progressive motility (PROG), Distance curved line (DCL), Distance average path (DAP), Distance straight line (DSL), Curvilinear velocity (VCL), Average path velocity (VAP), Straight-line velocity (VSL), Linearity of the curvilinear trajectory (LIN), Straightness (STR), the wobble of the curvilinear trajectory (WOB), the amplitude of lateral head displacement (ALH), and beat cross frequency (BCF).

Group	PROG (%)	DCL (μm)	DAP (μm)	DSL (μm)	VCL (μm/s)	VAP (μm/s)	VSL (μm/s)	LIN (%)	STR (%)	WOB (%)	ALH (μm)	BCF (Hz)
NT	51.43 ± 1.96	31.60 ± 0.69	29.24 ± 0.64	26.53 ± 0.65	67.70 ± 1.54	61.32 ± 1.56	55.91 ± 1.55	0.81 ± 0.004	0.91 ± 0.004	0.89 ± 0.005	1.01 ± 0.03	30.11 ± 0.38
GH +	38.11 ± 1.40	28.20 ± 1.02	25.72 ± 0.97	22.69 ± 0.99	59.26 ± 2.07	54.63 ± 2.15	47.88 ± 1.98	0.79 ± 0.004	0.88 ± 0.004	0.89 ± 0.005	0.89 ± 0.01	28.46 ± 0.60
$P < 0.05$	*	*	*	*							*	*

Asterisk indicates a significant difference between groups ($p < 0.05$). Data are expressed as means ± standard error of the mean (SEM) (n = 10 for each experimental group). NT, non-transgenic; gh+, gh-transgenic.

target gene present in TLR signaling pathway was *pik3cb*. As previously demonstrated, introducing a germline point mutation in *pik3cb* gene resulted in oligo- or azoospermia phenotypes in mice (Guillermet-Guibert et al., 2015). Taken together, these data suggest that this molecular pathway is closely related to sperm motility in different species, the result of an energetic trade-off between reproduction and immune responses. The negative regulation of miRNAs affects the genes downstream of the p53 signaling pathway in *gh* + zebrafish spermatozoa, indicating the activation of this pathway. These targets include those that induce cell-cycle arrest and initiate apoptosis by regulating several pro-apoptotic genes in another pathway that may be affected by the down-regulated miRNAs. As previously discussed, some events that occur during apoptosis, such as decreased mitochondrial functionality, spermatic membrane integrity, and DNA integrity, have also been observed in *gh* + zebrafish and appear to be involved in reducing fertilization and hatching frequency in these species (Figueiredo et al., 2013). In addition to the activation of the FoxO signaling pathway; therefore, the miRNAs also favor the expression of these pro-apoptotic genes (Wang et al., 2014).

Further, to comprehensively explore the biomarker potential of these miRNAs, a sensitive, accurate, and cost-efficient miRNA profiling technique is required. Next-generation sequencing (NGS) is emerging as a preferred method for miRNA profiling; as it offers high sensitivity, single-nucleotide resolution, and the possibility to profile a considerable number of samples in parallel (Coenen-Stass et al., 2018). Taking the above advantages into consideration, Vaz et al. (2015) demonstrated miRNA expression levels by miRNA sequencing of adult zebrafish tissues (brain, gut, liver, ovary, testis, eye, and heart) and compared with qPCR results showed good correlation for both known miRNAs and unknown miRNAs categories.

In addition to the above-mentioned study, identification and expression profile of miRNAs in spermatic cells and gonad of distinct species, such as zebrafish (Presslauer et al., 2017), Nile tilapia *Oreochromis niloticus* (Tao et al., 2016), *Bos taurus* (Stowe et al., 2014; Capra et al., 2017), *Sus scrofa* (Godia et al., 2019), and human (Hua et al., 2019; Xu et al., 2020), were done using NGS, without the need for further validation by qPCR.

The *in silico* analysis of pathways affected by the regulatory activity of down-regulated miRNAs identified by our research has risen some significant queries that need to be further examined, most notably the inference of miRNA sequencing data in F0104 males, a suitable translational model for studying the unpredictable collateral effects of excess GH on reproductive traits especially on the pathophysiology of male infertility by unknown, apparent cause. Follow-up studies may include experimental validation of functional interactions of miRNAs-target genes, using methods such as the luciferase assay report, which are undoubtedly required to understand the epigenetic mechanisms by which *gh* overexpression could contribute to a reduced reproductive potential.

In our study, in addition to the epigenetic modulation effect on spermatic cells from *gh* + zebrafish, we show the F0104 transgenic strain overexpressing GH showed a bodyweight 1.2

TABLE 2 | List of differentially expressed microRNAs (miRNAs) in sperm cells of *gh*-transgenic and non-transgenic zebrafish.

	miRNAs	Relative expression	Adjusted *P*-value	#Target genes
Down-regulated in *gh*+	dre-miR-19d-3p	0.11	0.002	202
	dre-miR-126a-5p	0.13	0.0005	392
	dre-miR-126b-5p	0.13	0.0005	392
	dre-miR-22a-5p	0.15	0.0001	46
	dre-miR-16c-5p	0.16	0.0002	187
	dre-miR-20a-5p	0.18	0.0005	404
	dre-miR-126b-3p	0.20	0.001	6
	dre-miR-107a-3p	0.20	0.001	106
	dre-miR-93	0.21	0.001	371
	dre-miR-2189	0.26	0.0003	142
	dre-miR-202-5p	0.27	0.00001	16
	dre-miR-221-3p	0.27	0.0002	69
	dre-miR-125a	0.30	0.002	49
	dre-miR-125b-5p	0.31	0.0003	51
	dre-miR-126a-3p	0.42	0.002	6
	dre-miR-30c-5p	0.43	0.004	374
Up-regulated in *gh*+	dre-miR-146b	4.99	0.000001	30
	dre-miR-200a-5p	6.38	0.003	26
	dre-miR-146a	6.90	0.000001	20
	dre-miR-726	8.31	0.0008	47
	dre-miR-184	17.44	0.00008	13
	dre-miR-738	70.59	0.000004	1

TABLE 3 | KEGG molecular pathways differentially expressed miRNA-targeted genes in sperm cells of *gh*-transgenic and non-transgenic zebrafish.

	KEGG molecular pathways	*P*-value	Target genes	#miRNA
Down-regulated miRNAs	FoxO signaling pathway	0.0006	26	11
	Protein processing in endoplasmic reticulum	0.004	22	9
	Apoptosis	0.01	14	8
	p53 signaling pathway	0.04	11	8
Up-regulated MiRNAs	Mucin type O-Glycan biosynthesis	0.00001	1	2
	Glycosaminoglycan biosynthesis	0.00002	1	1
	N-Glycan biosynthesis	0.002	1	1
	Toll-like receptor signaling pathway	0.01	2	2
	Glycerophospholipid metabolism	0.04	1	1

times greater than that of NT animals, confirming that the increased growth hormone led to stimulation of zebrafish muscle growth. Studies previously carried out using the same transgenic zebrafish strain reported similar muscle growth due to high levels of GH and thereby IGF1 expression, highly affecting the gene expressions essential for increased growth (Rosa et al., 2008; Kuradomi et al., 2011; Figueiredo et al., 2013; Silva et al., 2015; Nornberg et al., 2016).

Previous studies have demonstrated the differential expression of miRNA to be strongly associated with certain male reproductive dysfunctions, especially the movements of spermatozoa. Kinetic parameters are one of the main characteristic features of spermatozoa important for fertilization. A decrease in total motility and duration of spermatic motility observed in *gh* + zebrafish was in accordance with the previous findings in the F0104 strain (Figueiredo et al., 2013). In addition to motility, spermatic cells in *gh* + fish show decreased mitochondrial activity, membrane integrity, and DNA stability as the most probable reasons for reduced fertilization and hatching frequency seen in these fish (Figueiredo et al., 2013). Further, a significant increase in metabolic rate, generation of reactive oxygen species (ROS) (Rosa et al., 2008), and decrease in gene expressions related to the antioxidant defense system (Batista et al., 2014) in *gh* + fish can lead to oxidative stress, that in turn may cause deleterious effects on mitochondrial activity and ATP synthesis thereby reducing the fertilizing potential of spermatozoa.

In summary, our results conclude that *gh* overexpression alters the microRNAome expression profile of spermatic cells in zebrafish. The identified predicted gene targets from the up- and down-regulated miRNAs provide scope for future studies that involve validation of these miRNA-targets interactions to use

these candidate miRNAs as fertility biomarkers across species. Further, the *gh*-transgenic zebrafish strain may be used as a suitable model in reproductive translational research.

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/, SRP318063.

ETHICS STATEMENT

The animal study was reviewed and approved by the Ethics Committee of the Federal University of Rio Grande (FURG), Brazil, under the code 23116.008403/2018–32.

AUTHOR CONTRIBUTIONS

VC, WD, LM, and TS conceived and planned the experiments. TS, MK, and LM generated and raised the animals. WD, EB, IA, AV, and CC carried out the laboratorial analyses (spermatic kinetic and molecular biology). GG, CR, and WD contributed to miRNA libraries preparation. LN, AS, and DP contributed to the interpretation of the results. WD took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

FUNDING

This study was supported by the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS-FAPESP PqG #19/2551-0000953-3) and was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) Finance Code 001 and AUXPE #2537/2018. WD, LN, and EB were individually supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. AV, CC, LM, DP, and VC were also individually supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico.

ACKNOWLEDGMENTS

The undergraduate students from Laboratório de Genômica Estrutural for helping with daily routine at the Lab.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.704778/full#supplementary-material

Supplementary Figure 1 | Qualitative analysis of RNA quality isolated from zebrafish sperm cells samples. Representative total RNA samples quality parameter analyzed by TapeStation 4200. RNA from sperm cells of *gh*-transgenic (A1, B1, and C1) zebrafish (*Danio rerio*) belonging to the F0104 strain and non-transgenic groups (D1, E1, and F1) are compared with electronic Ladder (EL).

REFERENCES

Acosta, I. B., Junior, A. S. V., e Silva, E. F., Cardoso, T. F., Caldas, J. S., Jardim, R. D., et al. (2016). Effects of exposure to cadmium in sperm cells of zebrafish, Danio rerio. *Toxicol. Rep.* 3, 696–700. doi: 10.1016/j.toxrep.2016.08.002

Almeida, D. V., Bianchini, A., and Marins, L. F. (2013a). Growth hormone overexpression generates an unfavorable phenotype in juvenile transgenic zebrafish under hypoxic conditions. *Gen. Comp. Endocrinol.* 194, 102–109. doi: 10.1016/j.ygcen.2013.08.017

Almeida, D. V., de Martinez Gaspar Martins, C., de Azevedo Figueiredo, M., Lanes, C. F. C., Bianchini, A., and Marins, L. F. (2013b). Growth hormone transgenesis affects osmoregulation and energy metabolism in zebrafish (Danio rerio). *Transgenic Res.* 22, 75–88. doi: 10.1007/s11248-012-9627-x

Alves, M. B. R., Arruda, R. P., de Batissaco, L., Garcia-Oliveros, L. N., Gonzaga, V. H. G., Nogueira, V. J. M., et al. (2021). Changes in miRNA levels of sperm and small extracellular vesicles of seminal plasma are associated with transient scrotal heat stress in bulls. *Theriogenology* 161, 26–40. doi: 10.1016/j.theriogenology.2020.11.015

Batista, C. R., Figueiredo, M. A., Almeida, D. V., Romano, L. A., and Marins, L. F. (2014). Impairment of the immune system in GH-overexpressing transgenic zebrafish (Danio rerio). *Fish Shellfish Immunol.* 36, 519–524. doi: 10.1016/j.fsi.2013.12.022

Berishvili, G., Cotta, H. D., Baroiller, J., Segner, H., and Reinecke, M. (2006). Differential expression of IGF-I mRNA and peptide in the male and female gonad during early development of a bony fish, the tilapia Oreochromis niloticus. *Gen Comp Endocrinol* 146, 204–210. doi: 10.1016/j.ygcen.2005.11.008

Blödorn, E. B., Domingues, W. B., Nunes, L. S., Komninou, E. R., Pinhal, D., and Campos, V. F. (2021). MicroRNA roles and their potential use as selection tool to cold tolerance of domesticated teleostean species: A systematic review. *Aquaculture* 540:736747. doi: 10.1016/j.aquaculture.2021.736747

Cao, M., Chen, J., Peng, W., Wang, Y., Liao, L., Li, Y., et al. (2014). General and Comparative Endocrinology Effects of growth hormone over-expression on reproduction in the common carp Cyprinus carpio L. *Gen. Comp. Endocrinol.* 195, 47–57. doi: 10.1016/j.ygcen.2013.10.011

Capra, E., Turri, F., Lazzari, B., Cremonesi, P., Gliozzi, T. M., Fojadelli, I., et al. (2017). Small RNA sequencing of cryopreserved semen from single bull revealed altered miRNAs and piRNAs expression between High- and Low-motile sperm populations. *BMC Genomics* 18:1–12. doi: 10.1186/s12864-016-3394-7

Catalano-Iniesta, L., Sánchez-Robledo, V., Iglesias-Osma, M. C., García-Barrado, M. J., Carretero-Hernández, M., Blanco, E. J., et al. (2019). Sequential testicular atrophy involves changes in cellular proliferation and apoptosis associated with variations in aromatase P450 expression levels in Irs-2-deficient mice. *J. Anat.* 234, 227–243. doi: 10.1111/joa.12917

Chen, J., Cao, M., Zhang, A., Shi, M., Tao, B., Li, Y., et al. (2018). Growth hormone overexpression disrupts reproductive status through actions on leptin. *Front. Endocrinol.* 9:1–10. doi: 10.3389/fendo.2018.00131

Cito, G., Coccia, M. E., Salvianti, F., Fucci, R., Picone, R., Giachini, C., et al. (2020). Blood plasma miR-20a-5p expression as a potential non-invasive diagnostic biomarker of male infertility: a pilot study. *Andrology* 8, 1256–1264. doi: 10.1111/andr.12816

Coenen-Stass, A. M. L., Magen, I., Brooks, T., Ben-Dov, I. Z., Greensmith, L., Hornstein, E., et al. (2018). Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol.* 15, 1133–1145. doi: 10.1080/15476286.2018.1514236

Corral-vazquez, C., Salas-huetos, A., Blanco, J., and Vidal, F. (2019). Sperm microRNA pairs: new perspectives in the search for male fertility biomarkers. *Fertil. Steril.* 112, 831–841. doi: 10.1016/j.fertnstert.2019.07.006

Creaser, C. W. (1934). The technic of handling the zebra fish (Brachydanio rerio) for the production of eggs which are favorable for embryological research

and are available at any specified time throughout the year. *Copeia* 1934:159. doi: 10.2307/1435845

da Rosa, C. E., Figueiredo, M. A., Lanes, C. F. C., Almeida, D. V., and Marins, L. F. (2011). Genotype-dependent gene expression profile of the antioxidant defense system (ADS) in the liver of a GH-transgenic zebrafish model. *Transgenic Res.* 20, 85–89. doi: 10.1007/s11248-010-9395-4

Devlin, R. H., Biagi, C. A., and Yesaki, T. Y. (2004). Growth, viability and genetic characteristics of GH transgenic coho salmon strains. *Aquaculture* 236, 607–632. doi: 10.1016/j.aquaculture.2004.02.026

Domingues, W. B., Blodorn, E. B., Martins, A. S. W., Dellagostin, E. N., Komninou, E. R., Hurtado, J. I., et al. (2020). Transfection of exogenous DNA complexed to cationic dendrimer induces alterations of bovine sperm microRNAome. *Theriogenology* 156, 11–19. doi: 10.1016/j.theriogenology.2020.06.025

dos Santos da Silva, L., Borges Domingues, W., Fagundes Barreto, B., da Silveira, Martins, A. W., Dellagostin, E. N., et al. (2021). Capillary electroporation affects the expression of miRNA-122-5p from bull sperm cells. *Gene* 768:145286. doi: 10.1016/j.gene.2020.145286

Figueiredo, M., de, A., Lanes, C. F. C., Almeida, D. V., and Marins, L. F. (2007a). Improving the production of transgenic fish germlines: in vivo evaluation of mosaicism in zebrafish (Danio rerio) using flourescent protein (GFP) and growth hormone cDNA transgene co-injection strategy. *Genet. Mol. Biol.* 30, 31–36. doi: 10.1590/S1415-47572007000100008

Figueiredo, M. A., Fernandes, R. V., Studzinski, A. L., Rosa, C. E., Corcini, C. D., Varela Junior, A. S., et al. (2013). GH overexpression decreases spermatic parameters and reproductive success in two-years-old transgenic zebrafish males. *Anim. Reprod. Sci.* 139, 162–167. doi: 10.1016/j.anireprosci.2013.03.012

Figueiredo, M. A., Lanes, C. F. C., Almeida, D. V., Proietti, M. C., and Marins, L. F. (2007b). The effect of GH overexpression on GHR and IGF-I gene regulation in different genotypes of GH-transgenic zebrafish. *Comp. Biochem. Physiol.* 2, 228–233. doi: 10.1016/j.cbd.2007.04.004

Friedman, R. C., Farh, K. K.-H., Burge, C. B., and Bartel, D. P. (2008). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105. doi: 10.1101/gr.082701.108

Fromm, B., Domanska, D., Høye, E., Ovchinnikov, V., Kang, W., Aparicio-Puerta, E., et al. (2020). MirGeneDB 2.0: the metazoan microRNA complement. *Nucleic Acids Res.* 48, D132–D141. doi: 10.1093/nar/gkz885

Fujita, Y., Mihara, T., Okazaki, T., Shitanaka, M., Kushino, R., Ikeda, C., et al. (2011). Toll-like receptors (TLR) 2 and 4 on human sperm recognize bacterial endotoxins and mediate apoptosis. *Hum. Reprod.* 26, 2799–2806. doi: 10.1093/humrep/der234

Gac, F., Le, Blaise, O., Fostier, A., Bail, P., Le Loir, M., et al. (1993). Growth hormone (GH) and reproduction: a review. *Fish Physiol. Biochem.* 11, 219–232.

Gao, Y., Wu, F., Ren, Y., Zhou, Z., Chen, N., Huang, Y., et al. (2020). MiRNAs expression profiling of bovine (*Bos taurus*) testes and effect of bta-mir-146b on proliferation and apoptosis in bovine male germline stem cells. *Int. J. Mol. Sci* 21:21113846. doi: 10.3390/ijms21113846

Godia, M., Estill, M., Castello, A., Balasch, S., Rodriguez-Gil, J. E., Krawetz, S. A., et al. (2019). A RNA-seq analysis to describe the boar sperm transcriptome and its seasonal changes. *Front. Genet.* 10:299. doi: 10.3389/fgene.2019.00299

Griffeth, R. J., Carretero, J., and Burks, D. J. (2013). Insulin receptor substrate 2 is required for testicular development. *PLoS One* 8:62103. doi: 10.1371/journal.pone.0062103

Guillermet-Guibert, J., Smith, L. B., Halet, G., Whitehead, M. A., Pearce, W., Rebourcet, D., et al. (2015). Novel Role for p110β PI 3-Kinase in Male Fertility through Regulation of Androgen Receptor Activity in Sertoli Cells. *PLoS Genet.* 11:e1005304. doi: 10.1371/journal.pgen.1005304

Hua, M., Liu, W., Chen, Y., Zhang, F., Xu, B., Liu, S., et al. (2019). Identification of small non-coding RNAs as sperm quality biomarkers for in vitro fertilization. *Cell Discov.* 5:9. doi: 10.1038/s41421-019-0087-9

Hull, K. L., and Harvey, S. (2002). GH as a co-gonadotropin: the relevance of correlative changes in GH secretion and reproductive state. *J. Endocrinol.* 172, 1–19.

Ingham, P. W. (2009). The power of the zebrafish for disease analysis. *Hum. Mol. Genet.* 18, 107–112. doi: 10.1093/hmg/ddp091

Jia, K.-T., Zhang, J., Jia, P., Zeng, L., Jin, Y., Yuan, Y., et al. (2015). Identification of MicroRNAs in Zebrafish Spermatozoa. *Zebrafish* 12, 387–397. doi: 10.1089/zeb.2015.1115

Johnson, S. L., Zellhuber-McMillan, S., Gillum, J., Dunleavy, J., Evans, J. P., Nakagawa, S., et al. (2018). Evidence that fertility trades off with early offspring fitness as males age. *Proc. R. Soc. B Biol. Sci.* 285:20172174. doi: 10.1098/rspb.2017.2174

Kinth, P., Mahesh, G., and Panwar, Y. (2013). Mapping of zebrafish research: a global outlook. *Zebrafish* 10, 510–517. doi: 10.1089/zeb.2012.0854

Kotaja, N. (2014). MicroRNAs and spermatogenesis. *Fertil. Steril.* 101, 1552–1562. doi: 10.1016/j.fertnstert.2014.04.025

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162. doi: 10.1093/nar/gky1141

Kuradomi, R. Y., Figueiredo, M. A., Lanes, C. F. C., da Rosa, C. E., Almeida, D. V., Maggioni, R., et al. (2011). GH overexpression causes muscle hypertrophy independent from local IGF-I in a zebrafish transgenic model. *Transgenic Res.* 20, 513–521. doi: 10.1007/s11248-010-9429-y

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20. doi: 10.1016/j.cell.2004.12.035

Li, L., Zhu, Y., Chen, T., Sun, J., Luo, J., Shu, G., et al. (2019). MiR-125b-2 knockout in testis is associated with targeting to the PAP gene, mitochondrial copy number, and impaired sperm quality. *Int. J. Mol. Sci.* 20:148. doi: 10.3390/ijms20010148

Miura, C., Shimizu, Y., Uehara, M., Ozaki, Y., Young, G., and Miura, T. (2011). Gh is produced by the testis of Japanese eel and stimulates proliferation of spermatogonia. *Reproduction* 142, 869–877. doi: 10.1530/REP-11-0203

Moreau, D. T. R., Conway, C., and Fleming, I. A. (2011). Reproductive performance of alternative male phenotypes of growth hormone transgenic Atlantic salmon (Salmo salar). *Evol. Appl.* 4, 736–748. doi: 10.1111/j.1752-4571.2011.00196.x

Nam, Y. K., Cho, Y. S., Cho, H. J., and Kim, D. S. (2002). Accelerated growth performance and stable germ-line transmission in androgenetically derived homozygous transgenic mud loach. *Misgurnus Mizolepis. Aquacu.* 209, 257–270.

Navarro-Costa, P. A., Molaro, A., Misra, C. S., Meiklejohn, C. D., and Ellis, P. J. (2020). Sex and suicide: the curious case of Toll-like receptors. *PLoS Biol.* 18:e3000663. doi: 10.1371/journal.pbio.3000663

Nornberg, B. F., Almeida, D. V., Figueiredo, M. A., and Marins, L. F. (2016). GH indirectly enhances the regeneration of transgenic zebrafish fins through IGF2a and IGF2b. *Transgenic Res.* 25, 743–749. doi: 10.1007/s11248-016-9957-1

Paraskevopoulou, M. D., Georgakilas, G., Kostoulas, N., Vlachos, I. S., Vergoulis, T., Reczko, M., et al. (2013). DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.* 41, W169–W173. doi: 10.1093/nar/gkt393

Presslauer, C., Tilahun Bizuayehu, T., Kopp, M., Fernandes, J. M. O., and Babiak, I. (2017). Dynamics of miRNA transcriptome during gonadal development of zebrafish. *Sci. Rep.* 7:43850. doi: 10.1038/srep43850

Pursel, V. G., and Johnson, L. A. (1975). Freezing of boar spermatozoa: fertilizing capacity with concentrated semen and a new thawing procedure. *J. Anim. Sci.* 40, 99–102. doi: 10.2527/jas1975.40199x

Rahbar, S., Pashaiasl, M., Ezzati, M., Ahmadi AsrBadr, Y., Mohammadi-Dehcheshmeh, M., Mohammadi, S. A., et al. (2020). MicroRNA-based regulatory circuit involved in sperm infertility. *Andrologia* 52, 1–8. doi: 10.1111/and.13453

Rahman, M., Mak, R., Ayad, H., Smith, A., and Maclean, N. (1998). Expression of a novel piscine growth hormone gene results in growth enhancement in transgenic tilapia (Oreochromis niloticus). *Transgenic Res.* 7, 357–369.

Rosa, C. E., Figueiredo, M. A., Lanes, C. F. C., Almeida, D. V., Monserrat, J. M., and Marins, L. F. (2008). Metabolic rate and reactive oxygen species production in different genotypes of GH-transgenic zebrafish. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 149, 209–214. doi: 10.1016/j.cbpb.2007.09.010

Ruzicka, L., Howe, D. G., Ramachandran, S., Toro, S., Van Slyke, C. E., Bradford, Y. M., et al. (2019). The zebrafish information network: new support for non-coding genes, richer gene ontology annotations and the alliance of genome resources. *Nucleic Acids Res.* 47, D867–D873. doi: 10.1093/nar/gky1090

Schneider, A., Victoria, B., Lopez, Y. N., Suchorska, W., Barczak, W., Sobecka, A., et al. (2018). Tissue and serum microRNA profile of oral squamous cell carcinoma patients. *Sci. Rep.* 8:18945. doi: 10.1038/s41598-017-18945-z

Silva, A. C. G., Almeida, D. V., Nornberg, B. F., Figueiredo, M. A., Romano, L. A., and Marins, L. F. (2015). Effects of double transgenesis of somatotrophic axis (GH/GHR) on skeletal muscle growth of zebrafish (*Danio rerio*). *Zebrafish* 12, 408–413. doi: 10.1089/zeb.2015.29001.sil

Silveira, T., Kütter, M. T., Martins, C. M. G., Marins, L. F., Boyle, R. T., Campos, V. F., et al. (2021). First record of clinostomum sp. (Digenea: Clinostomidae) in danio rerio (Actinopterygii: Cyprinidae) and the implication of using zebrafish from pet stores on research. *Zebrafish* 18:1950.

Smorag, L., Zheng, Y., Nolte, J., Zechner, U., Engel, W., and Pantakani, D. V. K. (2012). MicroRNA signature in various cell types of mouse spermatogenesis: evidence for stage-specifically expressed miRNA-221, -203 and -34b-5p mediated spermatogenesis regulation. *Biol. Cell* 104, 677–692. doi: 10.1111/boc.201200014

Stowe, H. M., Calcatera, S. M., Dimmick, M. A., Andrae, J. G., Duckett, S. K., and Pratt, S. L. (2014). The bull sperm micrornaome and the effect of fescue toxicosis on sperm microRNA expression. *PLoS One* 9:e113163. doi: 10.1371/journal.pone.0113163

Streisinger, G., Walker, C., Dower, N., Knauber, D., and Singer, F. (1981). Production of clones of homozygous diploid zebra fish (Brachydanio rerio). *Nature* 291, 293–296. doi: 10.1038/291293a0

Studzinski, A. L. M., Barros, D. M., and Marins, L. F. (2015). Growth hormone (GH) increases cognition and expression of ionotropic glutamate receptors (AMPA and NMDA) in transgenic zebrafish (Danio rerio). *Behav. Brain Res.* 294, 36–42. doi: 10.1016/j.bbr.2015.07.054

Tao, M., Zhou, Y., Li, S., Zhong, H., Hu, H., Yuan, L., et al. (2018). MicroRNA alternations in the testes related to the sterility of triploid fish. *Mar. Biotechnol. N Y N* 20, 739–749. doi: 10.1007/s10126-018-9845-1

Tao, W., Sun, L., Shi, H., Cheng, Y., Jiang, D., Fu, B., et al. (2016). Integrated analysis of miRNA and mRNA expression profiles in tilapia gonads at an early stage of sex differentiation. *BMC Genomics* 17:328. doi: 10.1186/s12864-016-2636-z

Teame, T., Zhang, Z., Ran, C., Zhang, H., Yang, Y., Ding, Q., et al. (2019). The use of zebrafish (Danio rerio) as biomedical models. *Anim. Front.* 9, 68–77. doi: 10.1093/af/vfz020

van den Bos, R., Flik, G., and Gorissen, M. (2020). Behavioral research in zebrafish (Danio rerio): strain as source of variation. *Behav. Neural Genet. Zebrafish* 245–262. doi: 10.1016/b978-0-12-817528-6.00015-2

Vargas, R. (2018). Anesthesiology, anesthetics and zebrafish (Danio Rerio). an animal model to perform basic biomedical research. *EC Anaesth.* 4, 202–213.

Vaz, C., Wee, C. W., Lee, G. P. S., Ingham, P. W., Tanavde, V., and Mathavan, S. (2015). Deep sequencing of small RNA facilitates tissue and sex associated microRNA discovery in zebrafish. *BMC Genomics* 16:950. doi: 10.1186/s12864-015-2135-7

Vlachos, I. S., Zagganas, K., Paraskevopoulou, M. D., Georgakilas, G., Karagkouni, D., Vergoulis, T., et al. (2015). DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.* 43, W460–W466. doi: 10.1093/nar/gkv403

Wainwright, E. N., Jorgensen, J. S., Kim, Y., Truong, V., Bagheri-Fam, S., Davidson, T., et al. (2013). SOX9 regulates microRNA miR-202-5p/3p expression during mouse testis differentiation. *Biol. Reprod.* 89, 1–12. doi: 10.1095/biolreprod.113.110155

Wang, J. I. N., Chen, J., Sen, S., and Al, W. E. T. (2015). MicroRNA as biomarkers and diagnostics. *J. Cell. Physiol.* 231, 25–30. doi: 10.1002/jcp.25056

Wang, Y., Zhou, Y., and Graves, D. T. (2014). FOXO transcription factors: Their clinical significance and regulation. *BioMed Res. Int* 2014:925350. doi: 10.1155/2014/925350

Xie, X., Lu, J., Kulbokas, E. J., Golub, T. R., Mootha, V., Lindblad-toh, K., et al. (2005). Systematic discovery of regulatory motifs in human promoters and 3′. UTRs by comparison of several mammals. *Nature* 434, 338–345.

Xiong, S., Ma, W., Jing, J., Zhang, J., Dan, C., Gui, J. F., et al. (2018). An miR-200 cluster on chromosome 23 regulates sperm motility in zebrafish. *Endocrinology* 159, 1982–1991. doi: 10.1210/en.2018-00015

Xu, H., Wang, X., Wang, Z., Li, J., Xu, Z., Miao, M., et al. (2020). MicroRNA expression profile analysis in sperm reveals hsa-mir-191 as an auspicious omen of in vitro fertilization. *BMC Genomics* 21:165. doi: 10.1186/s12864-020-6570-8

Zhang, J., Ma, W., Xie, B., Gui, J.-F., and Mei, J. (2017a). Beneficial effect and potential molecular mechanism of chloroquine on sperm motility and fertilizing ability in yellow catfish. *Aquaculture* 468, 307–313. doi: 10.1016/j.aquaculture.2016.10.028

Zhang, J., Zhang, X., Liu, Y., Su, Z., Dawar, F. U., Dan, H., et al. (2017b). Leucine mediates autophagosome-lysosome fusion and improves sperm motility by activating the PI3K/Akt pathway. *Oncotarget* 8, 111807–111818. doi: 10.18632/oncotarget.22910

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 in Genetics

ORIGINAL RESEARCH
published: 20 September 2021
doi: 10.3389/fgene.2021.737906

Genomic Selection and Genome-wide Association Study for Feed-Efficiency Traits in a Farmed Nile Tilapia (*Oreochromis niloticus*) Population

Agustin Barría[1], *John A. H. Benzie*[2,3], *Ross D. Houston*[1], *Dirk-Jan De Koning*[4] *and Hugues de Verdal*[5,6,7,8]*

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh Easter Bush, Midlothian, United Kingdom, [2]WorldFish, Bayan Lepas, Malaysia, [3]School of Biological Earth and Environmental Sciences, University College Cork, Cork, Ireland, [4]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, [5]CIRAD, UMR ISEM, Montpellier, France, [6]ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France, [7]CIRAD, UMR AGAP Institut, Montpellier, France, [8]UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

OPEN ACCESS

Edited by:
Alexandre Wagner Silva Hilsdorf, University of Mogi das Cruzes, Brazil

Reviewed by:
Yi Zhou, Hunan Normal University, China
Yulin Jin, Emory University, United States

***Correspondence:**
Hugues de Verdal
hugues.de_verdal@cirad.fr

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

Received: *07 July 2021*
Accepted: *31 August 2021*
Published: *20 September 2021*

Citation:
*Barría A, Benzie JAH, Houston RD, 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*

Nile tilapia is a key aquaculture species with one of the highest production volumes globally. Genetic improvement of feed efficiency via selective breeding is an important goal, and genomic selection may expedite this process. The aims of this study were to *1*) dissect the genetic architecture of feed-efficiency traits in a Nile tilapia breeding population, *2*) map the genomic regions associated with these traits and identify candidate genes, *3*) evaluate the accuracy of breeding value prediction using genomic data, and *4*) assess the impact of the genetic marker density on genomic prediction accuracies. Using an experimental video recording trial, feed conversion ratio (FCR), body weight gain (BWG), residual feed intake (RFI) and feed intake (FI) traits were recorded in 40 full-sibling families from the GIFT (Genetically Improved Farmed *Tilapia*) Nile tilapia breeding population. Fish were genotyped with a ThermoFisher Axiom 65 K Nile tilapia SNP array. Significant heritabilities, ranging from 0.12 to 0.22, were estimated for all the assessed traits using the genomic relationship matrix. A negative but favourable genetic correlation was found between BWG and the feed-efficiency related traits; –0.60 and –0.63 for FCR and RFI, respectively. While the genome-wide association analyses suggested a polygenic genetic architecture for all the measured traits, there were significant QTL identified for BWG and FI on chromosomes seven and five respectively. Candidate genes previously found to be associated with feed-efficiency traits were located in these QTL regions, including *ntrk3a*, *ghrh* and *eif4e3*. The accuracy of breeding value prediction using the genomic data was up to 34% higher than using pedigree records. A SNP density of approximately 5,000 SNPs was sufficient to achieve similar prediction accuracy as the full genotype data set. Our results highlight the potential of genomic selection to improve feed efficiency traits in Nile tilapia breeding programmes.

Keywords: nile tilapia, feed conversion ratio, genomic prediction, SNP array, GWAS, candidate genes, feed intake

INTRODUCTION

One of the key features of aquaculture species compared with terrestrial farmed species is their greater feed efficiency (FE) (Brown, 2006). For example, fish need around six times less feed than cattle to produce the same amount of body mass (de Verdal et al., 2018a). However, feed still remains the primary cost for farmed fish production, and relatively little direct selection for improved feed efficiency has yet been performed for most aquaculture species. Therefore, genetic improvement of FE would enhance the economic sustainability of aquaculture and reduce environmental impacts, including greenhouse gas emissions (Aubin et al., 2009; MacLeod et al., 2020). Since domestication of aquaculture species is relatively young (Houston et al., 2020), selective breeding offers a great potential to improve commercially important traits (FAO, 2019). The benefits of genetic improvement have been illustrated in some key aquaculture species in relation to growth rate and disease resistance, particularly when augmented by genomic tools (see detailed reviews by Gjedrem and Rye, 2018; Fraslin et al., 2020; Houston et al., 2020). However, comparative little direct focus has been placed on feed efficiency in most aquaculture species, most likely due to the challenges of measuring feed intake efficiently and accurately at individual level (de Verdal et al., 2018a).

Nile tilapia (*Oreochromis niloticus*) is the third most produced farmed species globally (FAO, 2020). Farmed across a wide range of production systems, this species is considered a critical protein source for human consumption in undeveloped and developing countries (Miao and Wang, 2020). Recently, a wide variety of genomic tools have been developed for Nile tilapia (reviewed in Yáñez et al., 2020). These genomic tools have facilitated studies to unravel the relationships among improved strains (Hamilton et al., 2020), to detect regions associated with important traits (Cáceres et al., 2019; Yoshida et al., 2019; Taslima et al., 2020; Barria et al., 2021), and to assess the accuracy of prediction of breeding values using genomic selection (GS) models, compared with pedigree-based models (Yoshida et al., 2019; Joshi et al., 2020). However, there is a lack of studies targeting identification of genomic regions associated with FE traits (i.e. those which involve recording of feed intake), or assessing the potential impact of using genomic approaches to improve these traits in a Nile tilapia breeding populations. Previous work has aimed to assess body weight related traits (e.g. harvest weight, head weight, body length, fillet yield), as has been reviewed in Yáñez et al. (2020). Therefore, the aims of this study were to *1*) dissect the genetic architecture of feed-efficiency traits in a Nile tilapia breeding population, *2*) map the genomic regions associated with these traits and identify candidate genes, *3*) evaluate the accuracy of prediction of breeding values using genomic data, and *4*) assess the impact of the SNP density on the genomic prediction accuracies. To the best of our knowledge, this is the first study to assess genomic prediction for feed-efficiency traits in Nile tilapia. Our results highlight the feasibility of including feed-efficiency traits into Nile tilapia breeding programs. Furthermore, they show the potential improvement in breeding value prediction using low density SNP panels compared to using pedigree methods, which can enhance genetic gain and improve production efficiency.

MATERIALS AND METHODS

Nile Tilapia Breeding Population

This study used tissue samples archived by de Verdal et al. (2018b) from their study of genetic parameters of FE related traits in GIFT tilapia. In brief, the fish used in this study were from the Genetically Improved Farmed *Tilapia* (GIFT) breeding program based in Jitra, Malaysia, and managed by WorldFish (Ponzoni et al., 2011) These fish had been selected for increased growth rate for 15 generations in Malaysia at the time the experiment was performed. A total of 40 full-sibling families produced by natural spawning between December 2014 and December 2015 were used for the experimental challenge. All fish were fin clipped (the clip was stored individually in analytical grade ethanol) to provide tissue for DNA analysis at the beginning of the feed intake measurement stage (see next section).

Feed Conversion Experimental Challenge

Full details of the experimental challenge are given in previous studies, which reported methodology development and quantitative genetic analysis of feed efficiency traits (de Verdal et al., 2017; 2018b). In summary, a total of 1,200 fish (30 fish/family) were used for the feed efficiency experiment. There were four batches assessed during a period lasting from June 2015 to April 2016. For each batch, each full-sibling family was randomly split and transferred into two 200 L aquaria, such that each aquarium had a total of 15 fish. These fish were kept in the aquarium throughout the experimental challenge. To measure the amount of feed consumed daily by each fish, all fish were tagged with a unique two colored T-bar tag (Avery Dennison tags, 25 mm) combination at the dorsal muscle. The experimental challenge consisted of four different stages; 1) adaptation, 2) fasting, 3) feeding and 4) feed intake (FI), lasting 15, 10, 17 and 7 days, respectively. The number of days selected for each stage is described as follows: *1*) Generally 1 week is sufficient for an adaptation period. However, this was increased to 2 weeks as a precaution because some fish could adapt faster than others. The 10 days of fasting 2) represents the longest period permitted while maintaining acceptable fish welfare, while fewer days would probably have a low impact in terms of weight loss. The longest period was ascribed to the feeding stage 3), which allows the fish to recover from the fasting period but also enables successful measurement of compensatory growth, ensuring that this does not impact on the measurements of the feeding and growth during the last stage. Previous results by de Verdal et al. (2017), showed that the repeatability of FI was over 95%, when at least 11 meals were measured. Based on this, we selected 13 meals (7 days) for the current FI period 4).

Body weight was measured at the beginning and end of each stage. Due to aggression between fish, only 1,029 fish were available at the beginning of the FI stage where fish were fed twice a day (07:00 and 13:00), pellet by pellet, and videos of each

meal were recorded. Fish received a total of 13 meals during this stage (on the first day fish were weighed in the morning and only received one meal). Uneaten pellets were removed at the end of each meal. The number of pellets consumed by each fish was counted through video analysis. Throughout the experiment, water temperature was fixed at 28°C ± 1°C, while the photoperiod was 12L:12D cycle. The phenotypic sex of each fish was registered by visual observation of the gonads at the end of the FI recording period. The fish were sufficiently developed to show sexual differentiation, but were not yet exhibiting any reproductive behaviour. Mortality was recorded daily throughout the challenge. Thirty two fish died during the FI stage and were not included in the analyses.

Trait Definitions

As a practical calculation of pellet weight, a total of 500 pellets (16.4 ± 1.76 mg) were weighed, and as the variation in weight per pellet was considered low enough, the average weight of each pellet was used. Thus, the FI in grams for each fish could then be calculated as the sum of pellets consumed during the 13 meals. Body weight gain (BWG) was calculated during the FI stage as the difference between the body weight at the end and at the beginning of this stage. The quotient between these two traits (FI/BWG) was used to estimate the feed conversion ratio (FCR). Finally, residual feed intake (RFI) was estimated as the difference between the amount of feed intake by each fish and the amount predicted (Koch et al., 1963).

Outliers were highlighted using the boxplot. stats function of the R package "stats" (R Development Core Team, 2018) and were not included in the analyses.

SNP Array Genotyping

Total DNA was extracted from fin clips and genotyped by Identigen (Dublin, Ireland) by using a 65 K Axiom® SNP array (Peñaloza et al., 2020). The raw genotype data were filtered using the Axiom analysis Suite v4.0.3.3. Samples with a dish quality control <0.82 and/or genotype calling rate <0.93 were excluded from further analyses, leaving 801 samples. A total of 53 K SNPs were categorized as PolyHighResolution, and therefore were retained for subsequent analyses. Further quality control using Plink software v1.09 (Purcell et al., 2007) was performed. Specifically, SNPs were excluded for the genomic analyses due to either; *1*) call rate <0.95, *2*) minor allele frequency (MAF) < 0.05 or *3*) deviation from Hardy-Weinberg Equilibrium ($p < 1 \times 10^{-6}$). Simultaneously, animals with a call rate <0.95 were also excluded. Finally, to excluded potentially duplicated samples, both of each pair of fish with a proportion of IBD higher than 0.7 were removed from further analyses. Thus, a total of 755 fish and 48,431 SNPs remained after these filters, representing the final data for the SNP array (**Table 1**.).

To assess the genetic structure within the Nile tilapia population, a principal component analysis (PCA) was performed through PLINK v1.09. The two main components were plotted along the two axes in R.

Prediction of Breeding Values

Only fish with phenotype and genotype data were used for the pedigree-based BLUP (PBLUP) model, to enable a fair comparison between pedigree and genomic prediction performance. To predict the Estimated Breeding Values (EBVs) and estimate the variance components, the following univariate linear model was used:

TABLE 1 | Summary statistics for all the phenotyped fish for each feed-efficiency related trait.

Traits[a]	FCR	BWG	RFI	FI
N	726	755	727	735
Min	0.48	0.76	−4.41	1.91
Mean	0.94	9.18	−0.40	8.30
Std Dev	0.20	2.98	1.49	2.28
Median	0.91	8.94	−0.40	8.16
Max	1.55	18.24	4.40	15.22

[a]*FCR: Feed Conversion ratio, BWG: body weight gain (in g); FI: feed intake (in g); RFI: residual feed intake (in g).*

$$y = \mu + \boldsymbol{X}\beta + \boldsymbol{Z}a + \boldsymbol{W}c + e \quad (1)$$

Where y is the vector of phenotypes, μ is the overall average of phenotypic records, β is the vector of fixed effects accounting for sex and batch, a is the vector of random additive genetic effects, c is the vector of random effect associated with the common environmental effects (aquarium effect), e is the vector of random residuals errors, whereas $\boldsymbol{X}$, $\boldsymbol{Z}$ and $\mathbf{W}$ are the incidences matrices for the fixed, genetic and environmental effects, respectively. A normal distribution was assumed for the random genetic, common, and residuals effects, with a mean of zero and the following variance:

$$var\begin{bmatrix} a \\ c \\ e \end{bmatrix} = \begin{bmatrix} \mathbf{A}\sigma_a^2 & 0 & 0 \\ 0 & \mathbf{I}\sigma_c^2 & 0 \\ 0 & 0 & \mathbf{I}\sigma_e^2 \end{bmatrix}$$

Where σ_a^2, $\sigma_c^2$and σ_e^2 are the additive genetic, common environmental, and residual variance, respectively, whereas $\mathbf{A}$ and $\mathbf{I}$ are the numerator relationship and identity matrix, respectively. For all the traits, heritability was estimated as follows:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_c^2 + \sigma_e^2}$$

As described by VanRaden (2008), for the genomic BLUP (GBLUP), the numerator $\mathbf{A}$ relationship matrix was replaced by the genomic $\mathbf{(G)}$ relationship matrix. In this case, the vector of random additive genetic polygenic effects g has a distribution ~ N $(0, \sigma_g^2)$.

The single-step GBLUP (ssGBLUP) model was the same as for PBLUP and GBLUP. However, a combined kinship matrix H (Aguilar et al., 2010) which included information from the A and G matrix (Wang et al., 2012), was used. Thus, using data from pedigree and genotype data, the inverse of this H matrix is:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

Where A^{-1} and G^{-1} are the inverse of the numerator and genomic relationship matrix, respectively, and A_{22}^{-1} is the

inverse of the A matrix, but only for the genotyped fish. The estimation of the genetic parameters and models were fitted using the BLUPF90 family program (Misztal et al., 2015).

Predictive Ability and Cross Validation

The ability to predict the breeding values for the traits using pedigree and genomic approaches was assessed using a fivefold cross-validation approach. To assess the predictive ability among the different models, a total of 100 replicates were used. For each replicate, 80% of fish were used as a training dataset and, for the remaining 20%, the phenotypes were masked and used as the validation dataset. For each model, two cross-validation approaches were applied:

1)Random animal: for each replicate, 80% of all the fish were randomly drawn independently of their family, and used as a training dataset. The remaining 20% of the samples were used as a validation dataset.
2)Random families: for each replicate, 80% of the families were randomly drawn and used as a training set. For the remaining 20% of the families, the phenotype was masked and used as a validation dataset. Thus, breeding value predictions were made in the validation set using phenotypic data from different families.

For each approach, prediction accuracy was calculated as follows:

$$Accuracy = \frac{r(y_1,\ y_2)}{h}$$

Where y_1 and y_2 represents the predicted EBV and the phenotype, respectively, r is the correlation between these both values, and h is the square root of the estimated heritability.

Significant differences in predicted accuracies among models were assessed for each trait using both approaches. A Shapiro-Wilk test was done to evaluate the normality of the distribution of the predicted accuracies. The "rstatix" R package (Kassambara, 2021), was used for a pairwise comparison of the marginal means among models. Finally, the obtained p values were adjusted by the number of independent tests.

SNP Densities

Since the relatively high cost of high density SNP array genotyping may present a barrier to routine genotyping of large number of animals within a population, the efficacy of reduced density SNP panels for predicting breeding values was assessed. A total of 23 subsets of random SNPs were selected from the high density SNP array. These densities included 0.1, 0.3, 0.5, 0.7, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80 and 90% of the entire SNP dataset. For each SNP density, 100 replicates of a random animal five-fold cross validation were generated. The GBLUP model fitted for each density is the same as that described for the full genotype data set, and the accuracy was predicted as detailed previously.

Genome-wide Association Analyses

To identify genomic regions associated with the FE traits, a similar model to that proposed for genomic predictions was applied. However, since random common environmental effect was found to be non-significant, it was not included in the models. The GWAS was performed using the *leaving-one-chromosome-out* (LOCO) approach from the GCTA software v.1.92.2 (Yang et al., 2011). Briefly, a suggestive and genome-wide (α = 1 and 0.05, respectively) Bonferroni significance threshold was used to assess the significance of individual SNPs associated with the assessed traits. The former was first proposed by Lander and Kruglyak (1995), suggesting that at least one false positive marker is expected under the null hypothesis. Thus, a SNP was considered significant at a suggestive or genome-wide level if surpassed the corresponding Bonferroni significance threshold. The SNPs which were not mapped to a specific chromosome of the Nile tilapia reference genome (Conte et al., 2019; Genbank accession number GCA_001858045.3) were assigned to an artificial chromosome "Oni24".

For the significant QTLs associated with the traits, candidate genes were identified within a 500 Kb window size flanking the associated markers (250 Kb upstream and downstream, respectively) and used as the basis for a literature search for relevant functions connected to feed intake and efficiency. This size was selected based on Cádiz et al., 2020, which used a similar window size for look candidate genes in three commercial Nile tilapia populations.

RESULTS

FCR Challenge

Summary statistics for the feed efficiency and growth traits calculated after the trial are shown in **Table 1**. In brief, the BW of some fish increased by as little as 0.76 g, while others gained up to 18.24 g. In case of the FCR, the mean value was 0.94 ± 0.20, ranging from 0.48 to 1.55. The average RFI and FI were −0.40 ± 1.49 g and 8.30 ± 2.28 g, respectively, ranging from −4.41 to 4.40 g and from 1.91 to 15.22 g, respectively.

Genetic Parameters

The additive genetic variance estimated through PBLUP models was higher compared with GBLUP and ssGBLUP. The PBLUP model also had the lowest estimates for the environmental and residual variance (**Table 2**).

Heritability estimates using PBLUP, GBLUP and ssGBLUP were significant, and low to moderate in magnitude (**Table 2**). In all cases, the estimates were higher when a PBLUP model was fitted (ranging from 0.21 to 0.52), while both genomic models had similar estimates for a given trait (ranging from 0.12 to 0.22). No significant differences were found for c^2 among models.

A high genetic correlation of 0.98 ± 0.03 between FCR and RFI was found, suggesting that both measurements are essentially representations of the same trait (**Table 3**). Negative but favourable genetic correlations were found between BW, and both FCR (−0.60 ± 0.16) and RFI (−0.63 ± 0.17). However, when pedigree-based data was used, these correlations were not significantly different from zero (**Supplementary Table S1**).

Genome-wide Association

Principal components analyses indicated that the main two components accounted together for 15.6% of the total genetic

TABLE 2 | Estimated additive genetic (σ_a^2), common (σ_c^2), and residual (σ_e^2) variance parameters, common environmental effects (c^2) and heritabilities (h^2) for feed conversion ratio (FCR), body weight gain (BWG), residual feed intake (RFI) and feed intake (FI). Standard error is shown inside brackets.

Trait	Model[a]	σ_a^2	σ_c^2	σ_e^2	c^2	h^2
FCR	PBLUP	7.7E-03	1.7E-03	2.8E-02	0.05 (0.03)	0.21 (0.09)
	GBLUP	4.4E-03	2.4E-03	2.9E-02	0.07 (0.04)	0.12 (0.06)
	ssGBLUP	4.5E-03	2.3E-03	2.9E-02	0.06 (0.04)	0.12 (0.06)
BWG	PBLUP	1.82	0.18	4.60	0.03 (0.03)	0.28 (0.10)
	GBLUP	1.41	0.32	4.78	0.05 (0.03)	0.22 (0.07)
	ssGBLUP	1.46	0.31	4.81	0.05 (0.03)	0.22 (0.07)
RFI	PBLUP	0.96	3.7E-02	1.16	0.02 (0.02)	0.45 (0.12)
	GBLUP	0.37	0.14	1.49	0.07 (0.04)	0.19 (0.07)
	ssGBLUP	0.39	0.13	1.50	0.06 (0.04)	0.19 (0.07)
FI	PBLUP	2.19	2.5E-02	2.02	6.0E-03 (0.02)	0.52 (0.14)
	GBLUP	0.61	0.26	2.91	0.07 (0.04)	0.16 (0.06)
	ssGBLUP	0.64	0.24	2.91	0.06 (0.04)	0.17 (0.07)

[a]*PBLUP: pedigree-based BLUP models; GBLUP: genomic-based BLUP models; ssGBLUP: single-step genomic-based BLUP models.*

TABLE 3 | Estimated genetic correlation with standard error in brackets estimated using GBLUP models among feed conversion ratio (FCR), body weight gain (BWG), residual feed intake (RFI) and feed intake (FI).

	FCR	BWG	RFI	FI
FCR	–	–	–	–
BWG	-0.60 (0.16)	–	–	–
RFI	0.98 (0.03)	–0.63 (0.17)	–	–
FI	0.24 (0.25)	0.61 (0.16)	0.21 (0.23)	–

variation (**Figure 1S**), with no a clear structure evident within the population. There was a suggestive QTL on *Oni5* associated with both FCR and RFI (**Figures 1A, C**). The same SNPs were associated with both traits, consistent with the high positive genetic correlation between them. A genome-wide significant QTL for BWG (**Figure 1B**) was identified in the same region of chromosome *Oni5*, supported by a total of 10 SNPs. These ten SNPs were located within a 2.1 Mb window (**Table 4**). Lastly, there was a single SNP significantly associated with FI on chromosome *Oni7* (**Figure 1D**). The summary statistics for all the significant SNPs found for BWG and FI together with their detailed locations and allelic information are shown in **Table 4**.

The genes within a 500 Kb window surrounding the most significant SNPs associated with these traits are summarized in **Table 5**. The *Eukaryotic translation initiation factor 4E* (*eif4e3*) and *growth hormone-releasing hormone* (*ghrh*) were found within the QTL for BWG, *eif4e3* being situated approximately 89 Kbp upstream from the most significant SNP (AX-317169164), whereas *ghrh* is ~2 Kb upstream of the SNP AX-317150877. In the case of FI, the neurotrophic tyrosine kinase receptor type 3 (*ntrk3*), hyaluronidase 4 (*hyal4*) and *Sus scrofa* sperm adhesion molecule 1 (*spam1*), genes were found within the QTL region.

Accuracy of Breeding Value Predictions

Two different cross-validation approaches were tested to estimate prediction accuracies; a random individual and a random family approach. For the former, higher accuracies were achieved for BWG and FI when genomic data were used, with an increase of 34% compared with PBLUP (**Figure 2A**). However, in case of BWG this increase is only observed when the A and G matrix were combined (ssGBLUP).

As expected, all the models resulted in lower prediction accuracies when a random family approach was used, compared with the random animal approach. Furthermore, two main trends were observed. Firstly, for FCR and BWG, there were no significant differences in prediction accuracies using any of the models incorporating genomic data (**Figure 2B**). Secondly, the highest accuracies were achieved using GBLUP, representing an increase in 33 and 26% for RFI and FI respectively over PBLUP, while ssGBLUP gave intermediate prediction accuracies for FI.

Impact of SNP Density on Predicted Accuracies

A decrease in the estimated heritabilities were observed with the low density panels (data not shown). Therefore, for the calculation of prediction accuracy at all SNP densities, a single heritability estimated with all the available markers was used, as is considered the most accurate estimate (Kriaridou et al., 2020). Only a slight decrease in the prediction accuracies was observed across all traits when up to 10% of the SNP density was used (5 K SNPs). When SNP density fell below 5K SNPs, the breeding value prediction accuracy was lower for all the traits (**Figure 3**).

DISCUSSION

The current study generated genome-wide SNP genotype data from tissue samples collected from a previous study in which genetic parameters for FE traits were estimated in a Nile tilapia breeding population (de Verdal et al., 2018b). Through quantitative genetic analyses and using pedigree data with information from up to 15 generations, significant genetic

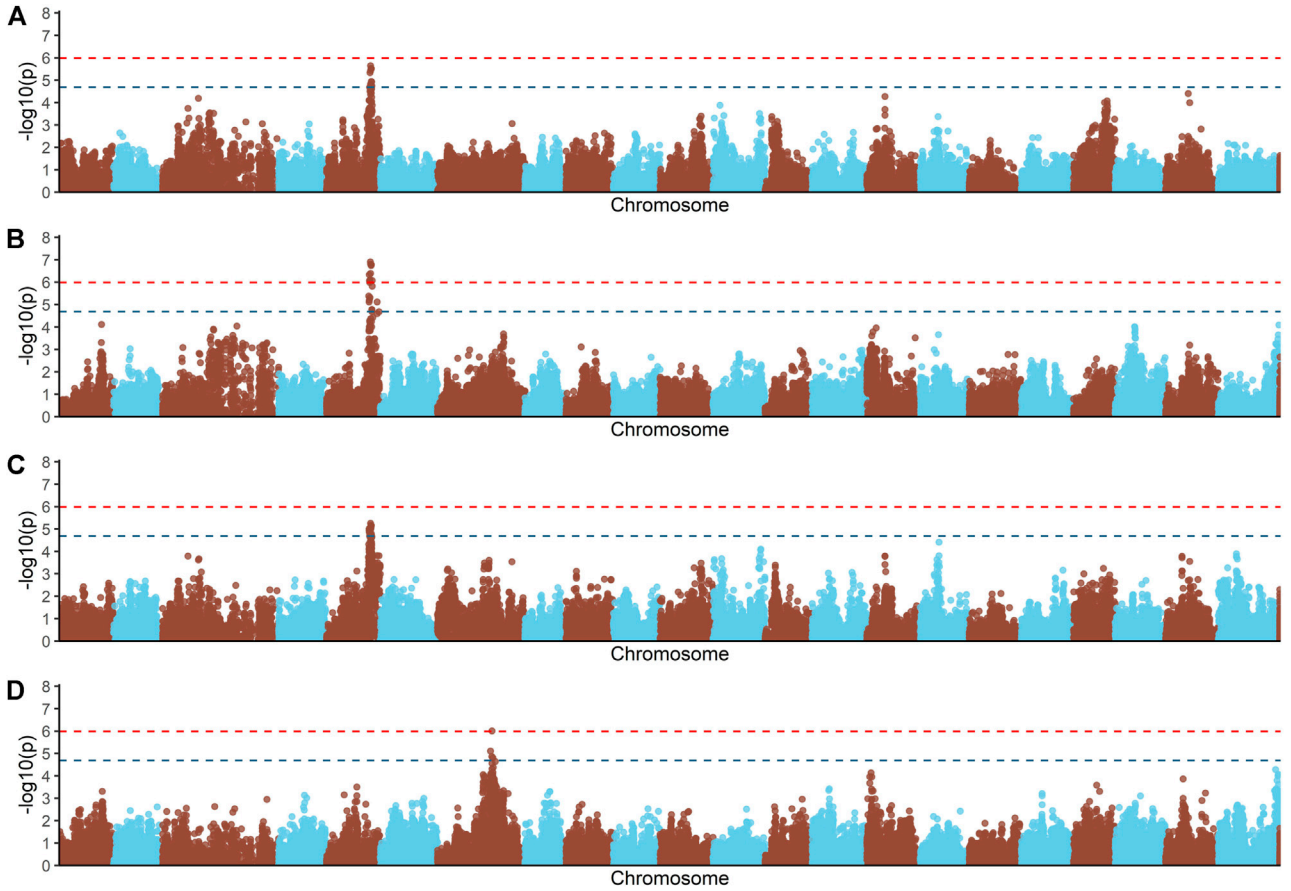

FIGURE 1 | Genome-wide association analysis for feed conversion rate related traits in a Nile tilapia (*Oreochromis niloticus*) breeding population. Genome-wide association analyses for feed conversion ratio **(A)**, body weight gain **(B)**, residual feed intake **(C)** and feed intake **(D)**. Red and blue dashed line represents the genome-wide and suggestive Bonferroni significant threshold, respectively.

TABLE 4 | Information on the genomic location, allelic variants and summary statistics for the genome-wide significant markers associated with feed-efficiency related traits in a Nile tilapia breeding population.

Body weight gain (BWG)					
***Oni*[a]**	**SNP**	**BP[b]**	**Minor allele**	**Major allele**	**Pval**
5	AX-317149233	31.78	A	C	8.7E-07
5	AX-317149515	32.14	G	A	7.8E-07
5	AX-317169164	32.44	C	T	1.2E-07
5	AX-317169167	32.47	G	A	1.6E-07
5	AX-317149783	32.53	C	T	4.1E-07
5	AX-317169548	32.99	T	C	1.8E-07
5	AX-317150097	32.99	A	G	1.7E-07
5	AX-317150157	33.02	G	A	1.7E-07
5	AX-317150877	33.69	A	G	8.5E-07
Feed intake (FI)					
7	AX-317210965	40.23	C	T	9.9E-07

[a]*Number of chromosome on the* Oreochromis niloticus *reference genome.*
[b]*Position of the SNP in the chromosome, in million base pairs.*

variation was detected for all the assessed traits, FCR, BWG, RFI and FI. The current study integrated the phenotypic dataset with genome-wide SNP information from recently developed SNP array for Nile *Tilapia* (Peñaloza et al., 2020), to estimate genetic correlations and genomic heritabilities using almost 50 K SNP markers segregating throughout the Nile tilapia genome. The same data were used to dissect the genetic architecture of feed-efficiency traits using a GWAS. Finally, the dataset was used to assess and compare genomic prediction accuracies using different statistical models, and assess the optimal SNP marker density for achieving maximal prediction accuracy.

Heritability and Genetic Correlations

Genetic parameters estimates for FE traits are relatively rare in aquaculture species, but have been extensively assessed in livestock species with significant heritabilites detected, ranging from 0.13 up to 0.84 (de Verdal et al., 2011; Case et al., 2012; Freetly et al., 2020; Tortereau et al., 2020; Marchesi et al., 2021). The studies on aquaculture species have also detected moderate to

TABLE 5 | Genes flanking the significant SNPs associated with feed-efficiency related traits in a Nile tilapia breeding population. Bold genes represent genes with potential impact on feed-efficiency related traits.

SNP	QTL region[b]		Gene names
	Left position	Right position	
Body weight gain (BWG)			
AX-317149064	31.31	31.81	*gnl3l, comt, traip, mon1a, trnar-acg,* ***bhlhe40, mst1r,*** *camkv, naaa, sema3f*
AX-317149233	31.53	32.03	*trnar-acg, bhlhe40,* ***trnt1,*** *camkv, naaa, sema3f*
AX-317149515	31.89	32.39	*trnt1, ddx4, gpr27, mdfic2, eif4e3, foxp1*
AX-317169164	32.19	32.69	*gpr27, ppp4r2, eif4e3, rybp, gxylt2, foxp1, pdzrn3, shq1*
AX-317169167	32.22	32.72	*gpr27, ppp4r2, eif4e3, rybp, gxylt2, foxp1, pdzrn3, shq1*
AX-317149783	32.28	32.78	*gpr27, ppp4r2,* ***eif4e3,*** *rybp, gxylt2, foxp1, pdzrn3, shq1*
AX-317169548	32.74	33.24	*chl1*
AX-317150097	32.74	33.24	*chl1*
AX-317150157	32.77	33.27	*chl1*
AX-317150877	33.44	33.94	*cdk5rap1,* ***ghrh,*** *ift52, srsf6,* ***rpn2,*** *mybl2,* ***acot8***
Feed Intake (FI)			
AX-317210965	39.98	40.48	*akap13, ska1, agbl1, ntrk3a, pard6a, wasla, hyal6, hyal4, spam1, gpr37a, pot1, grm8b, acot4*

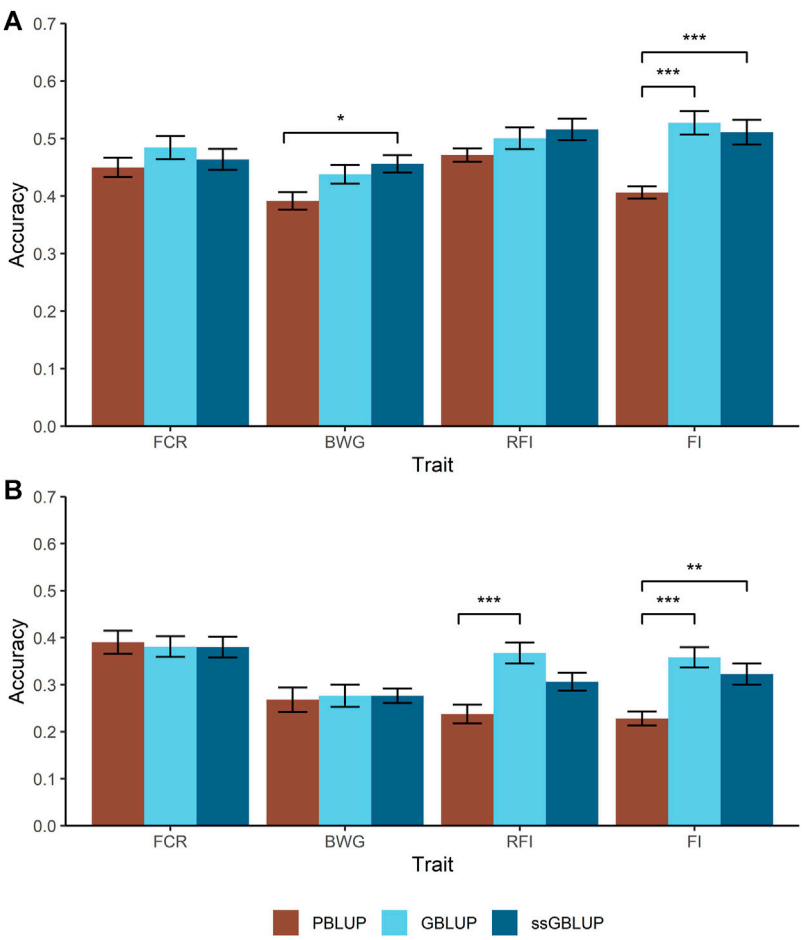

FIGURE 2 | Predicted accuracies comparison for the feed-efficiency related traits in a Nile tilapia (*Oreochromis niloticus*) breeding population. Accuracies predicted through pedigree-based BLUP (PBLUP), genomic BLUP (GBLUP) and single-step GBLUP (ssGBLUP) for feed conversion ratio (FCR), body weight gain (BWG), residual feed intake (RFI) and feed intake (FI) using a random animal **(A)** and random family **(B)** cross-validation approach. The standard error for each trait and method is represented by the black bars. * = $p < 0.05$; ** = $p < 1 \times 10^{-3}$; *** = $p < 1 \times 10^{-5}$.

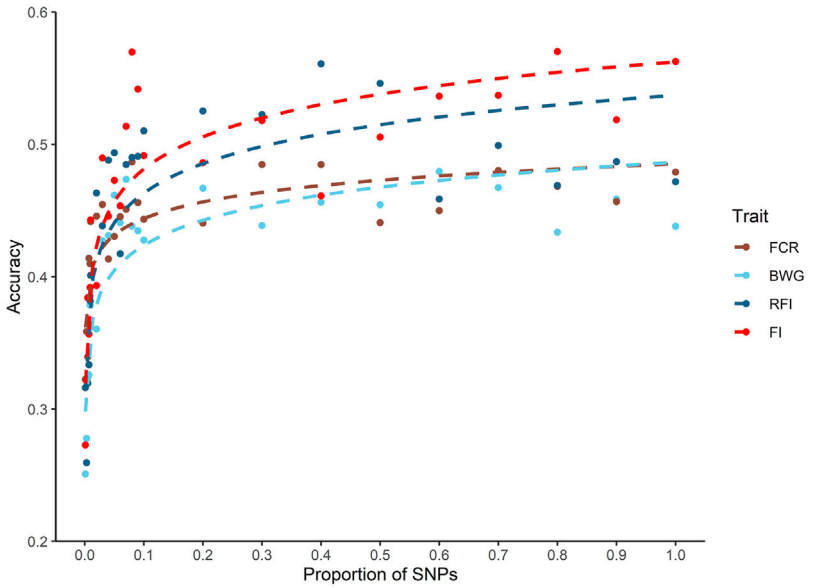

FIGURE 3 | Genomic prediction accuracies using different SNP densities for feed-efficiency related traits in a Nile tilapia (*Oreochromis niloticus*) breeding population. Predicted accuracies achieved using different SNP low density panels for feed conversion ratio (FCR), body weight gain (BWG), residual feed intake (RFI) and feed intake (FI) using a genomic BLUP (GBLUP) method.

high heritabilities (0.57–0.71) for average daily gain (ADG), daily feed intake (DFI), feed efficiency ratio (FER) and RFI in a breeding Pacific white shrimp (*Litopenaeus vannamei*) population (Dai et al., 2017). Furthermore, Kause et al. (2006) found significant heritabilities for growth in rainbow trout (*Oncorhynchus mykiss*), but not for FI, whereas Dvergedal et al. (2019) found significant heritabilities ranging from 0.18 to 0.45 for several feed efficiency traits in Atlantic salmon, such as weight gain, residual weight gain, and stable isotope profile in food. Furthermore, de Verdal et al. (2018b) estimated significant heritability (0.32–0.65) for FE traits in the current Nile tilapia population. Slightly lower heritabilities for the same traits were found in the current study, which could be explained by the ~25% fewer fish being included in the analyses.

A moderate genetic correlation was found between BWG and both FCR and RFI, when the genomic relationship matrix was used. This favourable and significant correlation suggests that selection for BWG (a simple trait to measure) would result in favourable correlated responses for feed efficiency traits within the current population. Although these results are in agreement with previous studies in sea bass (*Dicentrarchus labrax*) and rainbow trout (Kause et al., 2006, 2016; Silverstein, 2006; Besson et al., 2019), they contrast with those estimated by de Verdal et al. (2018b) who found no significant correlations. This cannot be attributed to differences in sample size, since correlations estimated using only pedigree data in the present study (**Supplementary Table S1**) were in accordance with de Verdal et al. (2018b). It is challenging to explain this discrepancy between the pedigree and the genomic data. It can be hypothesized that using genomic data, the true relationship between fish within each family were estimated in the genomic relationship matrix, exploiting two different levels of genetic variation: at inter and intra-family level. The latter is not included in the pedigree-based analyses. Therefore, using PBLUP, part of the additive effect went to the common environmental effect since families were not mixed in each aquarium, and so it was only possible to assess the differences between the families. In case of the genomic analyses, the genetic relationships between fish are more accurately estimated by estimating the random recombination during meiosis.

Genome-wide Association and Candidate Genes

With few exceptions (Houston et al., 2008; Moen et al., 2009; Boison et al., 2019; Sinclair-Waters et al., 2020; Barria et al., 2021), commercially important traits in aquaculture are underpinned by a polygenic genetic architecture (Tsai et al., 2015; Barría et al., 2018; Gutierrez et al., 2018; Palaiokostas et al., 2018; Mohamed et al., 2019; Aslam et al., 2020; Lu et al., 2020). This genetic architecture is also typical of FE traits in several breeding populations of livestock (Yuan et al., 2015; Higgins et al., 2018; Fu et al., 2020; Marchesi et al., 2021). Gutierrez et al. (2015) and Dvergedal et al. (2020) also found a polygenic architecture for these traits in Atlantic salmon. Whereas in the Crucian carp (*Carassius auratus*), several QTLs were detected for ADG, feed conversion efficiency and feed intake (Pang et al., 2017). Yoshida and Yáñez (2021) observed a polygenic architecture for ADG using a multi-trait GWAS approach in a Nile tilapia breeding population. Our results are in agreement with these previous studies, suggesting a polygenic architecture for feed efficiency traits in this current GIFT Nile tilapia breeding population. No major QTL were detected in the current study, but several QTLs were detected which explained a minor

proportion of genetic variation for the traits. Also, the possibility that some regions impacting the traits have not been covered by our genetic markers cannot be discarded.

The understanding of the interrelationship process among FE traits at the molecular level could provide new insights into the biological pathways involved and their genetic regulation. To date, a wide range of pathways have been associated with these traits, including forebrain development and neuron differentiation, hormone and growth factors, gluconeogenesis, lipogenesis, epithelial cell differentiation and hematopoietic cell lineage (McKenna et al., 2018; Li et al., 2019; Fu et al., 2020; Lindholm-Perry et al., 2020; Marchesi et al., 2021).

Among the genes flanking the significant markers located on *Oni5* and associated with BWG, *ghrh* was highlighted as one of the most important candidate genes given its strong association with growth rate through stimulation of growth hormone (GH). Fish and higher vertebrates growth is partly controlled by the GH/insulin-like growth factor-I (IGF-I) axis, which is also involved in the regulation of several physiological process in fish such as lipid and protein metabolism, immune function and feeding behaviour (Albalat et al., 2005; Kawauchi and Sower, 2006). For example, the modulation of genes related with the GH/IGF-I axis helps the fish to cope with fasting periods, by diverting the energy otherwise used for growth to other essential physiological processes (Wood et al., 2005; Gabillard et al., 2006). Although the relationship between growth and IGF-I is affected by several endogenous and exogenous factors, varying across seasons and productive cycles (Beckman, 2011), reliable associations have been shown in Mediterranean fishes as sea bass and gilthead sea bream (Company et al., 2001). In the case of tilapia, this association has also been found (Vera Cruz et al., 2006), where a GH-overexpressed transgenic line of Nile tilapia showed significantly higher weight and length than the wildtype (Eppler et al., 2007). Furthermore, in *O. mossambicus,* a recovery in plasma IGF-I levels during feeding stage after fasting, was associated with an increase in weight gain (Fox et al., 2009), and an administration of IGF-I and IGF-II has been shown to stimulate the growth rate (Chen et al., 2000).

Another interesting candidate gene, *eif4e3,* was located close to the most significant SNP for BWG. This gene belongs to the EIF4E family of proteins, which is associated with cell growth through protein synthesis and immune response (Piron et al., 2010). Previous work by Sun et al. (2016) showed significant differential expression of this gene, in *Hulong groupe* (*Epinephelus fuscogutatus x Epinephelus anceolatus*), a hybrid fish with an increased growth rate compared with their parental species.

It's well known the role of the brain-gut axis on the food intake regulation and nutrient metabolism (Reyer et al., 2015). For example, in zebrafish, modulation of brain activity through feeding and food availability has been reported, as well as an increase in feed intake after injection of a neurotrophic factor (Blanco et al., 2020). The gene *Neurotrophic Tyrosine Kinase Receptor Type 3*, *ntrk3*, the ligand of *neurotrophin* 3 (*ntf3*) may act as a modulator of feeding and satiety in mice and is associated with eating disorders in humans (Mercader et al., 2008). In the current study, this gene is located ~200 Kb from the significant SNPs for FI in *Oni07*, potentially playing a role on the amount of feed intake within the Nile tilapia population. In agreement with a study in another farmed species, *spam1* and *hyal4*, both genes found within the significant QTL for FI, have been previously reported as associated with FI in Landrace pigs. The authors suggest a role of these genes in fat synthesis and in lipid associated pathways, such as transport and metabolism (Fu et al., 2020). This could be explained by the fact that the amount of lipids plays a key role in the control of feed intake, as has been reported in livestock species (Harvatine and Allen, 2006; Kuhla et al., 2016), mainly by acting on the hypothalamic signals regulating feed consumption and energy expenditure (Allen et al., 2005; Relling et al., 2010).

Genomic Prediction and Impact of SNP Density

Since several factors such as heritability, linkage disequilibrium, population size and genetic architecture underlying traits can impact the accuracy of the genomic predictions (Morgante et al., 2018), it is necessary to compare the performance of different models when a trait is analyzed for the first time within a specific population. To our knowledge this is the first study to assess genomic predictions for feed-efficiency traits in Nile tilapia. Based on the GWAS results, the models used to predict the EBV accuracies were assumed to have a genetic variance controlled by thousands of markers with small effect (infinitesimal model).

The fact that there was no significant differences between the results for the GBLUP and ssGBLUP models for any of the traits and cross-validation approaches is likely due to the fact that only genotyped fish were included into the models. Therefore, the increase in the estimated relationship among fish due to the A and A_{22} matrices (and hence difference in the predicted accuracy of the EBV) is minimal, and most of the relationship is captured by the G matrix.

The observed differences in the EBV accuracies between the two cross-validation approaches (random versus family approaches) are likely to be explained by the genetic distance between fish from the training and validation groups. It has been shown that predicting EBV using data from fish from different full-sib families (i.e. random-family approach) affects mainly the pedigree-based model and traits that are not possible to measure on the candidate themselves (Tsai et al., 2016; Palaiokostas et al., 2019; Joshi et al., 2020). However, by including genomic data it is possible to exploit both inter- and intra-family genetic variation, achieving more accurate relationship assessments between individuals and therefore higher EBV accuracies (Garcia et al., 2018; Palaiokostas et al., 2018; Vallejo et al., 2019; Yoshida et al., 2019). Therefore, within a breeding programme, it would be preferable to ensure that measurements of FE traits were taken on training populations including close relatives, including full siblings, or the selection candidates. It should also be noted that FE measurements could potentially be taken on candidates themselves, but may be challenging for reasons of practicality and cost.

The impact of SNP density on predicted accuracies has been studied thoroughly for a wide variety of traits and species in aquaculture (Yoshida et al., 2019; Kriaridou et al., 2020; Tsairidou et al., 2020; Al-Tobasei et al., 2021; Vu et al., 2021). The results of these studies highlight the potential to significantly reduce SNP density without affecting the accuracy of prediction. Interestingly, Kriaridou et al. (2020) found an optimal SNP density of approximately 2 K SNPs for genomic prediction accuracy in different traits and species, suggesting a low impact of the genome size, genetic architecture, and family structure. The current study highlights that on average approximately 5 K is the minimal SNP density required to achieve similar accuracies to that of the 50 K SNP array for the tilapia population investigated. Thus, breeding programs could potentially use low density SNP markers (typically cheaper than high density panels) to achieve higher accuracies than using pedigree-based models, potentially increasing genetic gain in a cost-efficient manner.

CONCLUSION

Significant heritabilities were confirmed for feed-efficiency traits in a Nile tilapia breeding population, highlighting that genetic improvement is feasible. A negative, but favourable, genetic correlation was detected between BWG and FCR using the genomic data, suggesting selection for BWG may indirectly improve FCR. The traits were polygenic, suggesting the genomic selection may be the most effective route to incorporating genotype data into selection decisions. Genomic prediction markedly outperformed pedigree-based prediction, and this was the case even for relatively low density SNP markers. This study therefore highlights the potential for genomic selection to improve feed efficiency traits in Nile tilapia breeding programmes.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found here: doi:10.6084/m9.figshare.16573118.

ETHICS STATEMENT

Data collection and sampling was performed as part of a non-profit selective breeding program run by WorldFish. The animals from this breeding population are managed in accordance with the Guiding Principles of the Animal Care, Welfare and Ethics Policy of WorldFish Center including the "3-Rs" rule.

AUTHOR CONTRIBUTIONS

HdV designed and performed the animal experiment; HdV and AB performed the analyses with the advice of RH, JB, and DD; AB wrote the manuscript. All the co-authors revised and approved the final version of the manuscript.

FUNDING

This work was undertaken as part of, and funded by, the CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish. The program is supported by contributors to the CGIAR Trust Fund. Support was also provided by the International Fund for Agricultural Development (IFAD) and the European Commission-IFAD Grant Number 2000001539. The authors gratefully acknowledge funding from Institute Strategic Programme funding from the United Kingdom Biotechnology and Biological Sciences Research Council to The Roslin Institute (BB/P013759/1 and BB/P013740/1).

ACKNOWLEDGMENTS

We thank Hooi Ling Khaw, Hoong Yip Yee and Khairul Rizal Abu Bakar for the supply of fish and maintenance of the water system.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.737906/full#supplementary-material

REFERENCES

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot Topic: A Unified Approach to Utilize Phenotypic, Full Pedigree, and Genomic Information for Genetic Evaluation of Holstein Final Score. *J. Dairy Sci.* 93, 743–752. doi:10.3168/jds.2009-2730

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, 1–11. doi:10.1186/s12864-021-07404-9

Albalat, A., Gómez-Requeni, P., Rojas, P., Médale, F., Kaushik, S., Vianen, G. J., et al. (2005). Nutritional and Hormonal Control of Lipolysis in Isolated Gilthead Seabream (*Sparus aurata*) Adipocytes. *Am. J. Physiology-Regulatory, Integr. Comp. Physiol.* 289, R259–R265. doi:10.1152/ajpregu.00574.2004

Allen, M. S., Bradford, B. J., and Harvatine, K. J. (2005). The Cow as a Model to Study Food Intake Regulation. *Annu. Rev. Nutr.* 25, 523–547. doi:10.1146/annurev.nutr.25.050304.092704

Aslam, M. L., Boison, S. A., Lillehammer, M., Norris, A., and Gjerde, B. (2020). Genome-wide Association Mapping and Accuracy of Predictions for Amoebic Gill Disease in Atlantic salmon (*Salmo salar*). *Sci. Rep.* 10, 6435–6439. doi:10.1038/s41598-020-63423-8

Aubin, J., Papatryphon, E., van der Werf, H. M. G., and Chatzifotis, S. (2009). Assessment of the Environmental Impact of Carnivorous Finfish Production Systems Using Life Cycle Assessment. *J. Clean. Prod.* 17, 354–361. doi:10.1016/j.jclepro.2008.08.008

Barría, A., Christensen, K. A., Yoshida, G. M., Correa, K., Jedlicki, A., Lhorente, J. P., et al. (2018). Genomic Predictions and Genome-wide Association Study of Resistance against Piscirickettsia Salmonis in Coho Salmon (*Oncorhynchus*

kisutch) Using ddRAD Sequencing. *G3 (Bethesda)* 8, 1183–1194. 200053.2018. doi:10.1534/g3.118.200053

Barría, A., Trịnh, T. Q., Mahmuddin, M., Peñaloza, C., Papadopoulou, A., Gervais, O., et al. (2021). A Major Quantitative Trait Locus Affecting Resistance to Tilapia Lake Virus in Farmed Nile tilapia (*Oreochromis niloticus*). *Heredity* 127, 334–343. doi:10.1038/s41437-021-00447-4

Beckman, B. R. (2011). Perspectives on Concordant and Discordant Relations between Insulin-like Growth Factor 1 (IGF1) and Growth in Fishes. *Gen. Comp. Endocrinol.* 170, 233–252. doi:10.1016/j.ygcen.2010.08.009

Besson, M., Allal, F., Chatain, B., Vergnet, A., Clota, F., and Vandeputte, M. (2019). Combining Individual Phenotypes of Feed Intake with Genomic Data to Improve Feed Efficiency in Sea Bass. *Front. Genet.* 10, 1–14. doi:10.3389/fgene.2019.00219

Blanco, A. M., Bertucci, J. I., Hatef, A., and Unniappan, S. (2020). Feeding and Food Availability Modulate Brain-Derived Neurotrophic Factor, an Orexigen with Metabolic Roles in Zebrafish. *Sci. Rep.* 10, 10727. doi:10.1038/s41598-020-67535-z

Boison, S., Ding, J., Leder, E., Gjerde, B., Bergtun, P. H., Norris, A., et al. (2019). QTLs Associated with Resistance to Cardiomyopathy Syndrome in Atlantic Salmon. *J. Hered.* 110, 727–737. doi:10.1093/jhered/esz042

Brown, L. R. (2006). *Plan B 2.0: Rescuing a Planet under Stress and a Civilization in Trouble*.

Cáceres, G., López, M. E., Cádiz, M. I., Yoshida, G. M., Jedlicki, A., Palma-Véjares, R., et al. (2019). Fine Mapping Using Whole-Genome Sequencing Confirms Anti-müllerian Hormone as a Major Gene for Sex Determination in Farmed Nile Tilapia (*Oreochromis niloticus* L.). *G3 (Bethesda, Md.* 9, 3213–3223. doi:10.1534/g3.119.400297

Cádiz, M. I., López, M. E., Díaz-Domínguez, D., Cáceres, G., Yoshida, G. M., Gomez-Uchida, D., et al. (2020). Whole Genome Re-sequencing Reveals Recent Signatures of Selection in Three Strains of Farmed Nile tilapia (*Oreochromis niloticus*). *Sci. Rep.* 10, 11514. doi:10.1038/s41598-020-68064-5

Case, L. A., Wood, B. J., and Miller, S. P. (2012). The Genetic Parameters of Feed Efficiency and its Component Traits in the turkey (*Meleagris gallopavo*). *Genet. Sel Evol.* 44, 2. doi:10.1186/1297-9686-44-2

Chen, J.-Y., Chen, J.-C., Chang, C.-Y., Shen, S.-C., Chen, M.-S., and Wu, J.-L. (2000). Expression of Recombinant tilapia Insulin-like Growth Factor-I and Stimulation of Juvenile tilapia Growth by Injection of Recombinant IGFs Polypeptides. *Aquaculture* 181, 347–360. doi:10.1016/S0044-8486(99)00239-2

Company, R., Astola, A., Pendón, C., Valdivia, M. M., and Pérez-Sánchez, J. (2001). Somatotropic Regulation of Fish Growth and Adiposity: Growth Hormone (GH) and Somatolactin (SL) Relationship. *Comp. Biochem. Physiol. C: Toxicol. Pharmacol.* 130, 435–445. doi:10.1016/S1532-0456(01)00269-1

Conte, M. A., Joshi, R., Moore, E. C., Nandamuri, S. P., Gammerdinger, W. J., Roberts, R. B., et al. (2019). Chromosome-scale Assemblies Reveal the Structural Evolution of African Cichlid Genomes. *GigaScience* 8, 1–20. doi:10.1093/gigascience/giz030

Dai, P., Luan, S., Lu, X., Luo, K., Meng, X., Cao, B., et al. (2017). Genetic Assessment of Residual Feed Intake as a Feed Efficiency Trait in the Pacific white Shrimp Litopenaeus Vannamei. *Genet. Sel Evol.* 49, 61. doi:10.1186/s12711-017-0334-1

de Verdal, H., Mekkawy, W., Lind, C. E., Vandeputte, M., Chatain, B., and Benzie, J. A. H. (2017). Measuring Individual Feed Efficiency and its Correlations with Performance Traits in Nile tilapia, *Oreochromis niloticus*. *Aquaculture* 468, 489–495. doi:10.1016/j.aquaculture.2016.11.015

de Verdal, H., Narcy, A., Bastianelli, D., Chapuis, H., Même, N., Urvoix, S., et al. (2011). Improving the Efficiency of Feed Utilization in Poultry by Selection. 1. Genetic Parameters of Anatomy of the Gastro-Intestinal Tract and Digestive Efficiency. *BMC Genet.* 12, 59. doi:10.1186/1471-2156-12-59

de Verdal, H., Vandeputte, M., Mekkawy, W., Chatain, B., and Benzie, J. A. H. (2018b). Quantifying the Genetic Parameters of Feed Efficiency in Juvenile Nile tilapia *Oreochromis niloticus*. *BMC Genet.* 19, 1–10. doi:10.1186/s12863-018-0691-y

Dvergedal, H., Harvey, T. N., Jin, Y., Ødegård, J., Grønvold, L., Sandve, S. R., et al. (2020). Genomic Regions and Signaling Pathways Associated with Indicator Traits for Feed Efficiency in Juvenile Atlantic salmon (*Salmo salar*). *Genet. Sel Evol.* 52, 66. doi:10.1186/s12711-020-00587-x

Dvergedal, H., Ødegård, J., Øverland, M., Mydland, L. T., and Klemetsdal, G. (2019). Selection for Feed Efficiency in Atlantic salmon Using Individual Indicator Traits Based on Stable Isotope Profiling. *Genet. Sel Evol.* 51, 13. doi:10.1186/s12711-019-0455-9

Eppler, E., Caelers, A., Shved, N., Hwang, G., Rahman, A. M., Maclean, N., et al. (2007). Insulin-like Growth Factor I (IGF-I) in a Growth-Enhanced Transgenic (GH-Overexpressing) Bony Fish, the tilapia (*Oreochromis niloticus*): Indication for a Higher Impact of Autocrine/paracrine Than of Endocrine IGF-I. *Transgenic Res.* 16, 479–489. doi:10.1007/s11248-007-9093-z

FAO (2019). "The State of the World's Aquatic Genetic Resources for Food and Agriculture," in *Commission on Genetic Resources for Food and AgricultureCommission on Genetic Resources for Food and Agriculture*.

FAO (2020). *The State of World Fisheries and Aquaculture 2020: Sustainability in Action*. Rome, Italy: FAO. doi:10.4060/ca9229en

Fox, B. K., Breves, J. P., Hirano, T., and Grau, E. G. (2009). Effects of Short- and Long-Term Fasting on Plasma and Stomach Ghrelin, and the Growth Hormone/insulin-like Growth Factor I axis in the tilapia, *Oreochromis mossambicus*. *Domest. Anim. Endocrinol.* 37, 1–11. doi:10.1016/j.domaniend.2009.01.001

Fraslin, C., Quillet, E., Rochat, T., Dechamp, N., Bernardet, J.-F., Collet, B., et al. (2020). Combining Multiple Approaches and Models to Dissect the Genetic Architecture of Resistance to Infections in Fish. *Front. Genet.* 11, 677. doi:10.3389/fgene.2020.00677

Freetly, H. C., Kuehn, L. A., Thallman, R. M., and Snelling, W. M. (2020). Heritability and Genetic Correlations of Feed Intake, Body Weight Gain, Residual Gain, and Residual Feed Intake of Beef Cattle as Heifers and Cows. *J. Anim. Sci.* 98, skz394. doi:10.1093/jas/skz394

Fu, L., Jiang, Y., Wang, C., Mei, M., Zhou, Z., Jiang, Y., et al. (2020). A Genome-wide Association Study on Feed Efficiency Related Traits in Landrace Pigs. *Front. Genet.* 11, 692. doi:10.3389/fgene.2020.00692

Gabillard, J.-C., Kamangar, B. B., and Montserrat, N. (2006). Coordinated Regulation of the GH/IGF System Genes during Refeeding in Rainbow trout (*Oncorhynchus mykiss*). *J. Endocrinol.* 191, 15–24. doi:10.1677/joe.1.06869

Garcia, A. L. S., Bosworth, B., Waldbieser, G., Misztal, I., Tsuruta, S., and Lourenco, D. A. L. (2018). Development of Genomic Predictions for Harvest and Carcass Weight in Channel Catfish. *Genet. Sel Evol.* 50, 1–12. doi:10.1186/s12711-018-0435-5

Gjedrem, T., and Rye, M. (2018). Selection Response in Fish and Shellfish: a Review. *Rev. Aquacult* 10, 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, Genomes, Genet.* 8, 1273–1280. doi:10.1534/g3.118.200113

Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., and Davidson, W. S. (2015). Genome-wide Association Study (GWAS) for Growth Rate and Age at Sexual Maturation in Atlantic salmon (*Salmo salar*). *PLoS One* 10, e0119730. doi:10.1371/journal.pone.0119730

Hamilton, M. G., Lind, C. E., Barman, B. K., Velasco, R. R., Danting, M. J. C., and Benzie, J. A. H. (2020). Distinguishing between Nile Tilapia Strains Using a Low-Density Single-Nucleotide Polymorphism Panel. *Front. Genet.* 11, 594722. doi:10.3389/fgene.2020.594722

Harvatine, K. J., and Allen, M. S. (2006). Effects of Fatty Acid Supplements on Feed Intake, and Feeding and Chewing Behavior of Lactating Dairy Cows. *J. Dairy Sci.* 89, 1104–1112. doi:10.3168/jds.S0022-0302(06)72178-6

Higgins, M. G., Fitzsimons, C., McClure, M. C., McKenna, C., Conroy, S., Kenny, D. A., et al. (2018). GWAS and eQTL Analysis Identifies a SNP Associated with Both Residual Feed Intake and GFRA2 Expression in Beef Cattle. *Sci. Rep.* 8, 14301. doi:10.1038/s41598-018-32374-6

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

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, 1109–1115. doi:10.1534/genetics.107.082974

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

Kassambara, A. (2021). Rstatix: Pipe-Friendly Framework for Basic Statistical Tests. Available at: https://CRAN.R-project.org/package=rstatix [Accessed June 17, 2021].

Kause, A., Kiessling, A., Martin, S. A. M., Houlihan, D., and Ruohonen, K. (2016). Genetic Improvement of Feed Conversion Ratio via Indirect Selection against Lipid Deposition in Farmed Rainbow trout (*Oncorhynchus mykiss* Walbaum). *Br. J. Nutr.* 116, 1656–1665. doi:10.1017/S0007114516003603

Kause, A., Tobin, D., Houlihan, D. F., Martin, S. A. M., Mäntysaari, E. A., Ritola, O., et al. (2006). Feed Efficiency of Rainbow trout Can Be Improved through Selection: Different Genetic Potential on Alternative Diets1. *J. Anim. Sci.* 84, 807–817. doi:10.2527/2006.844807x

Kawauchi, H., and Sower, S. A. (2006). The Dawn and Evolution of Hormones in the Adenohypophysis. *Gen. Comp. Endocrinol.* 148, 3–14. doi:10.1016/j.ygcen.2005.10.011

Koch, R. M., Swiger, L. A., Chambers, D., and Gregory, K. E. (1963). Efficiency of Feed Use in Beef Cattle. *J. Anim. Sci.* 22, 486–494. doi:10.2527/jas1963.222486x

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, 1–8. doi:10.3389/fgene.2020.00124

Kuhla, B., Metges, C. C., and Hammon, H. M. (2016). Endogenous and Dietary Lipids Influencing Feed Intake and Energy Metabolism of Periparturient Dairy Cows. *Domest. Anim. Endocrinol.* 56, S2–S10. doi:10.1016/j.domaniend.2015.12.002

Lander, E., and Kruglyak, L. (1995). Genetic Dissection of Complex Traits: Guidelines for Interpreting and Reporting Linkage Results. *Nat. Genet.* 11, 241–247. doi:10.1038/ng1195-241

Li, B., Fang, L., Null, D. J., Hutchison, J. L., Connor, E. E., VanRaden, P. M., et al. (2019). High-density Genome-wide Association Study for Residual Feed Intake in Holstein Dairy Cattle. *J. Dairy Sci.* 102, 11067–11080. doi:10.3168/jds.2019-16645

Lindholm-Perry, A. K., Freetly, H. C., Oliver, W. T., Rempel, L. A., and Keel, B. N. (2020). Genes Associated with Body Weight Gain and Feed Intake Identified by Meta-Analysis of the Mesenteric Fat from Crossbred Beef Steers. *PLoS One* 15, e0227154. doi:10.1371/journal.pone.0227154

Lu, S., Liu, Y., Yu, X., Li, Y., Yang, Y., Wei, M., et al. (2020). Prediction of Genomic Breeding Values Based on Pre-selected SNPs Using ssGBLUP, WssGBLUP and BayesB for Edwardsiellosis Resistance in Japanese Flounder. *Genet. Sel Evol.* 52, 1–10. doi:10.1186/s12711-020-00566-2

MacLeod, M. J., Hasan, M. R., Robb, D. H. F., and Mamun-Ur-Rashid, M. (2020). Quantifying Greenhouse Gas Emissions from Global Aquaculture. *Sci. Rep.* 10, 11679. doi:10.1038/s41598-020-68231-8

Marchesi, J. A. P., Ono, R. K., Cantão, M. E., Ibelli, A. M. G., Peixoto, J. d. O., Moreira, G. C. M., et al. (2021). Exploring the Genetic Architecture of Feed Efficiency Traits in Chickens. *Sci. Rep.* 11, 4622. doi:10.1038/s41598-021-84125-9

McKenna, C., Porter, R. K., Keogh, K. A., Waters, S. M., McGee, M., and Kenny, D. A. (2018). Residual Feed Intake Phenotype and Gender Affect the Expression of Key Genes of the Lipogenesis Pathway in Subcutaneous Adipose Tissue of Beef Cattle. *J. Anim. Sci Biotechnol* 9, 68. doi:10.1186/s40104-018-0282-9

Mercader, J. M., Saus, E., Agüera, Z., Bayés, M., Boni, C., Carreras, A., et al. (2008). Association of NTRK3 and its Interaction with NGF Suggest an Altered Cross-Regulation of the Neurotrophin Signaling Pathway in Eating Disorders. *Hum. Mol. Genet.* 17, 1234–1244. doi:10.1093/hmg/ddn013

Miao, W., and Wang, W. (2020). Trends of Aquaculture Production and Trade: Carp, tilapia, and Shrimp. *Afs* 33S, 1–10. doi:10.33997/j.afs.2020.33.S1.001

Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., Legarra, A., and Vitezica, Z. (2015). Manual for BLUPF90 Family of Programs. 125. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf.

Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a Major QTL for Resistance to Infectious Pancreatic Necrosis in Atlantic salmon (*Salmo salar*): Population-Level Associations between Markers and Trait. *BMC genomics* 10, 368. doi:10.1186/1471-2164-10-368

Mohamed, A. R., Verbyla, K. L., Al-Mamun, H. A., McWilliam, S., Evans, B., King, H., et al. (2019). Polygenic and Sex Specific Architecture for Two Maturation Traits in Farmed Atlantic salmon. *BMC Genomics* 20, 1–13. doi:10.1186/s12864-019-5525-4

Morgante, F., Huang, W., Maltecca, C., and Mackay, T. F. C. (2018). Effect of Genetic Architecture on the Prediction Accuracy of Quantitative Traits in Samples of Unrelated Individuals. *Heredity* 120, 500–514. doi:10.1038/s41437-017-0043-0

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, 30–11. doi:10.1186/s12711-018-0401-2

Palaiokostas, C., Vesely, T., Kocour, M., Prchal, M., Pokorova, D., Piackova, V., et al. (2019). Optimizing Genomic Prediction of Host Resistance to Koi Herpesvirus Disease in Carp. *Front. Genet.* 10, 1–9. doi:10.3389/fgene.2019.00543

Pang, M., Fu, B., Yu, X., Liu, H., Wang, X., Yin, Z., et al. (2017). Quantitative Trait Loci Mapping for Feed Conversion Efficiency in Crucian Carp (*Carassius auratus*). *Sci. Rep.* 7, 1–11. doi:10.1038/s41598-017-17269-2

Peñaloza, C., Robledo, D., Barría, A., Trịnh, T. Q., Mahmuddin, M., Wiener, P., et al. (2020). Development and Validation of an Open Access SNP Array for Nile tilapia (*Oreochromis niloticus*). *G3 Genes Genomes Genet.* 10, 2777–2785. doi:10.1534/g3.120.401343

Piron, F., Nicolaï, M., Minoïa, S., Piednoir, E., Moretti, A., Salgues, A., et al. (2010). An Induced Mutation in Tomato eiF4E Leads to Immunity to Two Potyviruses. *PLoS ONE* 5, e11313. doi:10.1371/journal.pone.0011313

Ponzoni, R. W., Nguyen, N. H., Khaw, H. L., Hamzah, A., Bakar, K. R. A., and Yee, H. Y. (2011). Genetic Improvement of Nile tilapia (*Oreochromis niloticus*) with Special Reference to the Work Conducted by the WorldFish Center with the GIFT Strain. *Rev. Aquac.* 3, 27–41. doi:10.1111/j.1753-5131.2010.01041.x3

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

R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at http://www.R-project.org/.

Relling, A. E., Pate, J. L., Reynolds, C. K., and Loerch, S. C. (2010). Effect of Feed Restriction and Supplemental Dietary Fat on Gut Peptide and Hypothalamic Neuropeptide Messenger Ribonucleic Acid Concentrations in Growing Wethers1. *J. Anim. Sci.* 88, 737–748. doi:10.2527/jas.2009-2316

Reyer, H., Hawken, R., Murani, E., Ponsuksili, S., and Wimmers, K. (2015). The Genetics of Feed Conversion Efficiency Traits in a Commercial Broiler Line. *Sci. Rep.* 5, 16387. doi:10.1038/srep16387

Silverstein, J. T. (2006). Relationships Among Feed Intake, Feed Efficiency, and Growth in Juvenile Rainbow Trout. *North Am. J. Aquac.* 68, 168–175. doi:10.1577/a05-010.1

Sinclair-Waters, M., Ødegård, J., Korsvoll, S. A., Moen, T., Lien, S., Primmer, C. R., et al. (2020). Beyond Large-Effect Loci: Large-Scale GWAS Reveals a Mixed Large-Effect and Polygenic Architecture for Age at Maturity of Atlantic salmon. *Genet. Sel Evol.* 52, 1–11. doi:10.1186/s12711-020-0529-8

Sun, Y., Guo, C.-Y., Wang, D.-D., Li, X. F., Xiao, L., Zhang, X., et al. (2016). Transcriptome Analysis Reveals the Molecular Mechanisms Underlying Growth Superiority in a Novel Grouper Hybrid (Epinephelus Fuscogutatus♀ × E. Lanceolatus♂). *BMC Genet.* 17, 1–10. doi:10.1186/s12863-016-0328-y

Taslima, K., Wehner, S., Taggart, J. B., De Verdal, H., Benzie, J. A. H., Bekaert, M., et al. (2020). Sex Determination in the GIFT Strain of tilapia Is Controlled by a Locus in Linkage Group 23. *BMC Genet.* 21, 1–15. doi:10.1186/s12863-020-00853-3

Tortereau, F., Marie-Etancelin, C., Weisbecker, J.-L., Marcon, D., Bouvier, F., Moreno-Romieux, C., et al. (2020). Genetic Parameters for Feed Efficiency in Romane Rams and Responses to Single-Generation Selection. *animal* 14, 681–687. doi:10.1017/S1751731119002544

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Gharbi, K., Stear, M. J., et al. (2015). Genome Wide Association and Genomic Prediction for Growth Traits in Juvenile Farmed Atlantic salmon Using a High Density SNP Array. *BMC Genomics* 16, 1–9. doi:10.1186/s12864-015-2117-9

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, 47–11. 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: Genes, Genomes, Genet.* 10, 581–590. doi:10.1534/g3.119.400800

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, 47–14. doi:10.1186/s12711-019-0489-z

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

Vera Cruz, E. M., Brown, C. L., Luckenbach, J. A., Picha, M. E., Bolivar, R. B., and Borski, R. J. (2006). Insulin-like Growth Factor-I cDNA Cloning, Gene Expression and Potential Use as a Growth Rate Indicator in Nile tilapia, *Oreochromis niloticus*. *Aquaculture* 251, 585–595. doi:10.1016/j.aquaculture.2005.06.039

Verdal, H., Komen, H., Quillet, E., Chatain, B., Allal, F., Benzie, J. A. H., et al. (2018a). Improving Feed Efficiency in Fish Using Selective Breeding: A Review. *Rev. Aquacult* 10, 833–851. doi:10.1111/raq.12202

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

Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. M. (2012). Genome-wide Association Mapping Including Phenotypes from Relatives without Genotypes. *Genet. Res.* 94, 73–83. doi:10.1017/S0016672312000274

Wood, A. W., Duan, C., and Bern, H. A. (2005). Insulin-like Growth Factor Signaling in Fish. *Int. Rev. Cytol.* 243, 215–285. doi:10.1016/S0074-7696(05)43004-1

Yáñez, J. M., Joshi, R., and Yoshida, G. M. (2020). Genomics to Accelerate Genetic Improvement in tilapia. *Anim. Genet.* 51, 658–674. doi:10.1111/age.12989

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, 76–82. doi:10.1016/j.ajhg.2010.11.011

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: Genes, Genomes, Genet.* 9, 2597–2607. doi:10.1534/g3.119.400116

Yoshida, G. M., and Yáñez, J. M. (2021). Multi-trait GWAS Using Imputed High-Density Genotypes from Whole-Genome Sequencing Identifies Genes Associated with Body Traits in Nile tilapia. *BMC Genomics* 22, 57. doi:10.1186/s12864-020-07341-z

Yuan, J., Wang, K., Yi, G., Ma, M., Dou, T., Sun, C., et al. (2015). Genome-wide Association Studies for Feed Intake and Efficiency in Two Laying Periods of Chickens. *Genet. Sel Evol.* 47, 82. doi:10.1186/s12711-015-0161-1

Conflict of Interest: JB was employed by the company WorldFish.

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 in Genetics

ORIGINAL RESEARCH
published: 01 November 2021
doi: 10.3389/fgene.2021.696893

Quantitative Genetics of Smoltification Status at the Time of Seawater Transfer in Atlantic Salmon (*Salmo Salar*)

Hooi Ling Khaw[1], *Bjarne Gjerde*[1]*, *Solomon A. Boison*[1,2], *Elise Hjelle*[3] *and Gareth F. Difford*[1]

[1]Department of Breeding and Genetics, Nofima AS, Osloveien, Norway, [2]Mowi ASA, Bergen, Norway, [3]PHARMAQ Analytiq AS, Bergen, Norway

High mortality during grow out in the sea is a challenge for farmed Atlantic salmon production in Norway and globally, which is partly attributed to suboptimal smolt quality. In this study, two groups of pre-smolts were put on a standard light smoltification regime with alternating 12L:12D per day for 6 weeks (Phase I), followed by 24L:0D per day for 6 weeks (Phase II); one group was 0 + smolt (EXP1) and the other 1 + smolt (EXP2). To monitor the smoltification status of the fish, 100 (EXP1) and 60 (EXP2) fish were randomly sampled per week during Phase II. The following phenotypes for smoltification status were studied: RT-qPCR relative mRNA expression of values of two alpha catalytic subunits of the variants of the Na^+K^+ATPase (NKA) expressed in the sampled gill tissues of each fish. The first variant, alpha1a with increased expression in freshwater (FW) and the second variant alpha1b with increased expression in seawater variant (SW), as well as their ratio SW/FW. At the optimal time for seawater transfer based on the SW/FW trait, 1,000 (at sixth sampling of EXP1) and 1,500 (at fifth sampling of EXP2) fish were sampled for genetic parameter estimation. The individual variation in FW, SW, and SW/FW was very large at each of the seven samplings indicating a large variation among individuals in the optimum time of transfer to seawater. SW/FW showed significant genetic variation in both 0+ and 1+ smolts, which indicates the possibility for selection for improved synchronization of smoltification status of Atlantic salmon at the time where the largest proportion of the fish is considered to be smolt. However, the genetic correlation between SW/FW of 0+ and 1+ was not significantly different from zero indicating very little shared genetic variation in SW/FW in 0+ and 1+ fish. Smoltification phenotypes showed temporal progression over the smoltification period, and this progression varied between 0+ and 1+ smolt highlighting the importance of correctly timing the major sampling point, and when cohorts are transferred to seawater. This also highlighted the need for further research into noninvasive methods of objectively measuring individual smoltification through time and subsequent smolt survival and growth rate at sea.

Keywords: smoltification, atlantic salmon, 0+ and 1+ smolts, heritability, genetic correlations, optimum seawater transfer

OPEN ACCESS

Edited by:
Qiang Lin,
South China Sea Institute of Oceanology (CAS), China

Reviewed by:
Roberto Neira,
University of Chile, Chile
Giri Athrey,
Texas A and M University, United States

***Correspondence:**
Bjarne Gjerde
bjarne.gjerde@nofima.no

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

Received: *18 April 2021*
Accepted: *27 September 2021*
Published: *01 November 2021*

Citation:
*Khaw HL, Gjerde B, Boison SA, Hjelle E and Difford GF (2021) Quantitative Genetics of Smoltification Status at the Time of Seawater Transfer in Atlantic Salmon (*Salmo Salar*). Front. Genet. 12:696893. doi: 10.3389/fgene.2021.696893*

INTRODUCTION

Atlantic salmon (*Salmo salar*) is the major aquaculture commodity in Norway, which in 2019 amounted to 1.357 million tons (94% of the total aquaculture production of Norway) to a value of NOK 68.1 billion [USD 7.0 billion (https://www.ssb.no)]. Mortality during grow out in the sea continues to remain a challenge for Atlantic salmon production in Norway as well as globally and represents a substantial economic loss for the farmers. Aunsmo et al. (2008) reported significant losses occurring during the first 2–3 months after seawater transfer and attributed this to suboptimal smolt production and quality, spurring intensive research into improved survival and welfare of post smolts in the seawater phase (Kolarevic et al., 2014; Ytrestøl et al., 2020).

Smoltification is the process where Atlantic salmon parr undergo behavioral, developmental, and physiological changes into smolt, which in wild fish enable their first migration downstream to the sea (Hoar, 1988) and in farmed fish to be transferred primarily into floating net cages in the sea. These changes include alterations to body shape, skin reflectance, increased sodium–potassium (Na^+/K^+) ATPase in gills, and the osmotic ability to regulate blood plasma ion concentrations at seawater phase (Stefansson et al., 2008). Although both wild and domesticated Atlantic salmon undergo smoltification, the timing, age, and size of the fish can be vastly different. In wild Atlantic salmon, parr are locked into smoltification once their body size and energy status reach an underlying threshold in the spring, and this can happen anytime between 1 and 8 years of age (Thorstad et al., 2012). The parr remain in the resting state until smoltification is triggered by a zeitgeber, in this case increasing photoperiod and water temperatures in the spring (Thorpe et al., 1998).

In the formative years of the Atlantic salmon industry, a population of parr would partially smoltify under natural water temperatures and photoperiods in the spring approximately 16 months after hatching and are commonly called (1+) smolts to indicate they are older than 1 year. The remaining parr, which failed to smoltify, would then continue to be cultured in freshwater until their second spring approximately 28 months after hatching and would be called (2+) smolt to indicate they are older than 2 years (Refstie et al., 1977). In the current domesticated Atlantic salmon, genetic selection for improved growth, improved feed formulation, more optimal water temperature, and rearing environments have enabled the majority of smolts to be produced at an age of 16 months after hatching (1+) and out of season smolts using artificial lighting and controlled water temperature at an age of 7–8 months after hatching. These are commonly referred to as 0+ smolts in keeping with the classification of 1+ and 2+ smolts (Mørkøre et al., 2001). An important point is that Atlantic salmon is not expressly selected for improved and synchronized 0+ or 1+ smoltification status, but rather for increased growth and there is little knowledge on whether different selection strategies are needed for these two smolt production strategies.

The physiology of seawater osmoregulation of salmonids is well understood; however, there is limited information on underlying causes of genetic variation in salinity tolerance of the fish. Currently, smoltification status of fish is determined by methods that could be classified into noninvasive and invasive. The noninvasive methods are through monitoring changes on body morphology, skin reflectance, and/or K-factor (condition factor), which tend to be less objective and accurate (Stefansson et al., 2003; Piironen et al., 2013). The invasive methods are all based on evaluating osmoregulatory ability, particularly the activity of the $Na^+/K+$ ATPase enzyme (NKA) in the gills. The NKA enzyme is comprised of two essential isoforms of the alpha catalytic subunit: the alpha 1a subunit associated with expression in freshwater and ionic uptake, and the alpha 1b subunit isoform expression associated with saltwater and ionic excretion (McCormick et al., 2013). The first method includes challenging fish with exposure to full-strength saltwater and then recording the blood plasma ion concentrations (Cl^-, Na^+, Mg^{2+}) to evaluate the performance of the NKA pathway in regulating blood ion concentrations (Stefansson et al., 2003; Piironen et al., 2013). The second method involves a gill biopsy, an enzymatic activity assay using spectrophotometry or antibodies against the NKA protein (McCormick et al., 1993; McCormick et al., 2013). Another method involves real-time quantitative polymerase chain reaction (RT-qPCR) measurements of the expression on the two isoforms of the NKA alpha subunit (McCormick et al., 2009; Handeland et al., 2014). Although the gill biopsies can be collected nonlethally (McCormick et al., 1993), in many countries including Norway, it is a legal requirement to sample fish lethally dosed with anesthesia postmortem (Stefansson et al., 2003; Piironen et al., 2013). This makes it impossible to follow the individual growth rate of fish sampled for these phenotypes after seawater transfer and also inhibits their use as breeding candidates for genetic improvement.

Phenotypes that could effectively and accurately quantify smoltification status will provide the needed information for studying the genetic architecture of smoltification. Even though several studies have reported results on mapping of quantitative trait loci (QTL) for smoltification status traits in salmonids (Boulding et al., 2008; Nichols et al., 2008; Stefansson et al., 2008; Le Bras et al., 2011), the information on genetic variation for smoltification status remains very limited. Thus, the objectives of the present study were to estimate the genetic variation for different traits related to smoltification status at the time of transfer from freshwater to seawater and the genetic correlations between these smoltification traits within and between a group of 0+ and a group of 1+ smolts partly from the same families.

MATERIALS AND METHODS

The Environment and the Fish

The smoltification experiment was conducted at the freshwater facility of Nofima Research Station for Sustainable Aquaculture (RSSA), located at Sunndalsøra (62°40′3.5292″N,

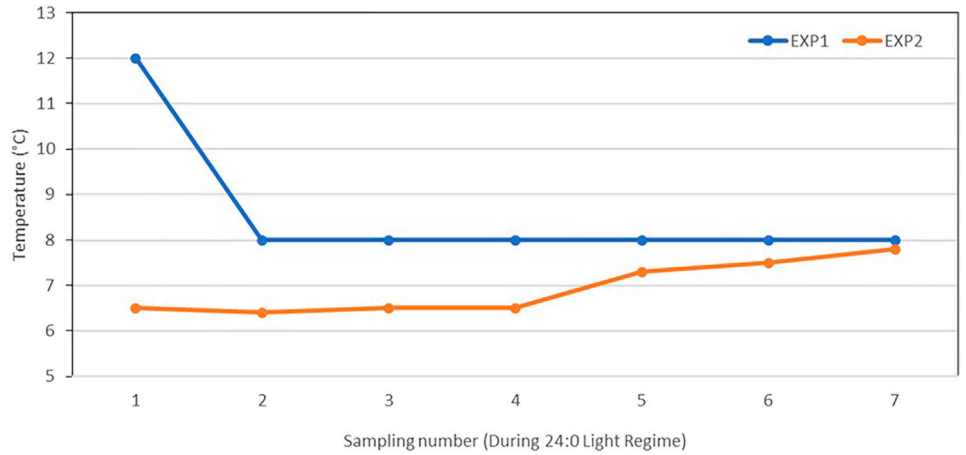

FIGURE 1 | Freshwater temperatures (°C) at each of the seven samplings during Phase II of 0+ smolt (EXP1) (6 weeks) and 1+ smolt (EXP2) (10 weeks).

8°31′28.974″E), Norway. In January 2017, Mowi ASA delivered around 3,000 eyed eggs of Atlantic salmon from 50 families (60 eggs per family, from 25 sires and 50 dams) to RSSA, where the eggs were hatched and reared in one tank to an average body size of 50 g, as the experimental fish (0+) for the first experiment (EXP1). In July 2017, a batch of 2,500 fingerlings of size 30 g from 150 families from 50 sires and 150 dams, which included the sibs of the 50 families that involved in EXP1, were sent from Mowi ASA to RSSA. The fingerlings were reared until they reached an average body weight of 200 g (1+) by the end of January 2018 and used in the second experiment (EXP2). All the fish were not individually identified with a PIT tag, but at the time of gill sampling and body measurement recording, a fin sample from the tail was taken for genotyping using a medium density SNP array (57K Affymetrix chip developed by Nofima) for the purpose of parentage assignment.

The Experimental Design and Fish

The smoltification experiments were conducted in two batches with the 0+ parr from September 12 to December 5, 2017 (EXP1) and 1+ parr from February 5 to May 31, 2018 (EXP2). For EXP1, 2,000 fish (40 fish from each of the 50 families) with an average body weight of 50 g at the start of the experiment were stocked in one experimental tank of size 3 m × 3 m × 0.8 m. For EXP2, two groups of 1,000 fish each (13–14 fish from each of 150 families) with an average body weight of 200 g at the start of the experiment were stocked in two experimental tanks (T1 and T2) each of size 3 m × 3 m × 0.8 m.

A standard light regime was used for smoltification, which started with alternating 12 h of light and 12 h of darkness (12L: 12D) for 6 weeks (Phase I), followed by continuous (24 h, 24L: 0D) light for 6 weeks (Phase II). For EXP2, Phase II lasted for 10 weeks and, thus, 4 weeks longer than planned. This was mainly due to the water temperatures at that time were lower (6.4°C–7.8°C) than during Phase I (**Figure 1**), and therefore, the fish needed more days to reach the desired degree-days (°d = °C × days) for smoltification [300 to 400°days after the increase in the daylength as suggested by Sigholt et al. (1998)].

TABLE 1 | Three characteristics for smolt index scoring.

Characteristic	Index (point)[a]			
Parr mark	Clear (1)	Visible (2)	Weak (3)	None (4)
Silver coloration	Clear (1)	Weak (2)	Visible (3)	Silver (4)
Fin margins	Clear (1)	Weak (2)	Visible (3)	Black margin (4)

Note:
[a]*The transition of Atlantic salmon parr to smolt is indicated in the gradual increasing score from 1 to 4.*

For monitoring and measuring the smoltification status of the fish, 100 (EXP1) and 60 fish (EXP2; 30 fish per tank) were randomly sampled during Phase II on a weekly basis for EXP1 and at slightly longer intervals of 10–11 days for EXP2 to account for the lower temperature in EXP2 (**Figure 1**). At the optimal time for seawater transfer (based on the results from SmoltVision analysis provided by Pharmaq Analytiq), 1,000 (EXP1) and 1,500 fish (EXP2; 750 fish per tank) were sampled for genetic parameter estimation. For EXP1, this major sampling event took place on November 28, 2017 (at the sixth of the seven samplings), and for EXP2, on May 14, 2018 (at the fifth of the seven samplings).

The sampled fish were first anesthetized with a lethal dose of tricaine mesylate (MS-222). Once anesthetized, a gill filament was biopsied from the left side of the fish and placed in the SmoltVision tube containing RNAlater (ThermoFisher, United States) for preservation of the sample, then sent to Pharmaq Analytiq for further analysis. Immediately after gill biopsy, the fish were given a sharp blow to the head to ensure no chance of unintended recovery, as stipulated in the ethical permits. The ethical use of fish in both experiments were approved by the Norwegian Food Safety Authority (Mattilsynet) with approval IDs 13174 (EXP1) and 15310 (EXP2).

Smoltification Phenotypes and Records

Four smoltification phenotypes were studied relating to the commercial SmoltVision test (Pharmaq Analytic, Norway): the RT-qPCR relative expression of mRNA encoding for the two protein isoforms of the NKA alpha catalytic subunit; the first one encoding for the NKA alpha 1a subunit, which is upregulated in freshwater and will be referred to as FW, and the second encoding for the NKA alpha 1b subunit, which is upregulated in saltwater and will be referred to as SW for the purpose of this study. The third phenotype is the ratio of the relative expression of the SW over FW variants referred to as SW/FW, as this is the primary quantity used in practice to determine whether fish are ready for seawater transfer. The fourth phenotype is the smolt index (SI), which is a composite phenotype calculated as the average of three subjectively scored external characteristics of the fish (parr marks, silver coloration, and fin edges, **Table 1**).

Importantly, the researcher must submit metadata accompanying the gill biopsy samples. This includes the tank or cage number, water temperature, light regime as well as the individual body length, body weight, and the three abovementioned external smolt characteristics (**Table 1**). Pharmaq Analytic returns the RT-qPCR relative expression of mRNA encoding for the two protein isoforms of the NKA alpha catalytic subunit and a second gene encoding for the $NA^+K^+Cl^-$ co-transporter (NK1CC) used as a positive internal control, expressed relative to the elongation factor 1alpha (EF1a) housekeeping gene. Importantly, the researcher has no control over the analysis and processing of these values, and these details are proprietary information and are restricted from the user. However, the SmoltVision test is based on the methods previously described by Handeland et al. (2014).

In addition, the individual body weight (BW) and body length (BL) were recorded for each fish sampled at each of the seven samplings. The condition factor was calculated as $K-\text{factor} = \frac{BW*100}{BL^b}$, where b = 3 as in the general formula of the K-factor (Nash et al., 2006), and K2-factor where the b was estimated for EXP1 (b = 2.845 ± 0.040) and EXP2 (b = 2.788 ± 0.019) by fitting a simple linear regression model, $\ln(BW) = \text{intercept} + b*\ln(BL) + e$. This was done to obtain a K2-factor that was independent of BL of the fish (Le Cren, 1951). The phenotypic correlation of K-factor with BL in EXP1 was −0.18 (p-value <0.0001) and in EXP2 −0.32 (p-value <0.0001), while that of K2-factor with BL in EXP1 was 0.046 (p = 0.13) and in EXP2 0.004 (p = 0.88).

Estimation of Phenotypic and Genetic Parameters

The data from EXP1 and EXP2 were first analyzed separately to obtain genetic parameters of the traits within the 1+ and 0+ smolts and, thereafter, combined primarily to obtain an estimate of the genetic correlation between the homologous trait measured in the 0+ and 1+ fish. All the parameter estimates were obtained using restricted maximum likelihood in WOMBAT (Meyer, 2007) and with the genomic relationship matrix added to the random animal effect of the mixed model equations of the applied linear mixed animal model (see below).

Parameter Estimates for the Different Traits Within EXP1 and EXP2

Phenotypic and genetic (co)variances for the studied traits, i.e., BW, K2-factor, FW, SW, and SW/FW ratio, were obtained from a multitrait animal model, separately for EXP1 and EXP2. In obtaining all the genetic correlations between these five traits, a three-trait (BW, K2-factor, SW/FW) and a four-trait (BW, K2-factor, FW, and SW) model was used as the parameters did not converge when all five traits were included simultaneously, specifically when combinations of SW/FW, SW, and FW were included. The four-trait model may be written as:

$$\begin{bmatrix}\mathbf{y}_1\\ \mathbf{y}_2\\ \mathbf{y}_3\\ \mathbf{y}_4\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & 0 & 0 & 0\\ 0 & \mathbf{X}_2 & 0 & 0\\ 0 & 0 & \mathbf{X}_3 & 0\\ 0 & 0 & 0 & \mathbf{X}_4\end{bmatrix}\begin{bmatrix}\mathbf{b}_1\\ \mathbf{b}_2\\ \mathbf{b}_3\\ \mathbf{b}_4\end{bmatrix} + \begin{bmatrix}\mathbf{Z}_{A_1} & 0 & 0 & 0\\ 0 & \mathbf{Z}_{A_2} & 0 & 0\\ 0 & 0 & \mathbf{Z}_{A_3} & 0\\ 0 & 0 & 0 & \mathbf{Z}_{A_4}\end{bmatrix} \times \begin{bmatrix}\mathbf{a}_1\\ \mathbf{a}_2\\ \mathbf{a}_3\\ \mathbf{a}_4\end{bmatrix} + \begin{bmatrix}\mathbf{e}_1\\ \mathbf{e}_2\\ \mathbf{e}_3\\ \mathbf{e}_4\end{bmatrix};$$

where for each of the traits, **y** is the vector of the phenotypic observations, **b** is the vector of fixed effect (overall mean, tank with two levels (T1 or T2) for EXP2, sex with the two levels male or female), and $\mathbf{X}$ and $\mathbf{Z}_A$ are the design matrices assigning observations to their levels of the fixed and animal additive genetic effects, respectively.

a is the vector of random additive genetic effects, which follows normal distribution $N(0, G\otimes A)$, where G is the marker-based genomic relationship matrix constructed as $\frac{MM'}{2\sum p_i(1-p_i)}$, M is the centered marker genotypes, and p_i is the allele frequency of each marker; A is the genetic variance covariances matrix; and **e** is the vector of random residual effects, which follows the normal distribution $N(0, I\otimes R)$, where I is the identity matrix, and R is the residual variance covariances matrix.

The trait BL was not included due to its high phenotypic and genetic correlations to BW (r_g = 0.94 ± 0.00; r_g = 0.97 ± 0.01 for Exp1; r_p = 0.96 ± 0.00, r_g = 0.97 ± 0.01 for EXP2) obtained from a bivariate analysis of the data from each of the two experiments. For the trait SI in EXP2, the genetic variation was zero and was therefore not included in the model. The effect of sex was not significantly different from zero for all the traits (p > 0.05) and was therefore omitted from the final model.

Genetic Correlation Between the Homologous Trait in EXP1 and EXP2

In obtaining the genetic correlation between the six homologous traits in EXP1 and EXP2, i.e., BW, K2-factor, SI, FW, SW, and SW/FW, each trait was considered as two different traits in a

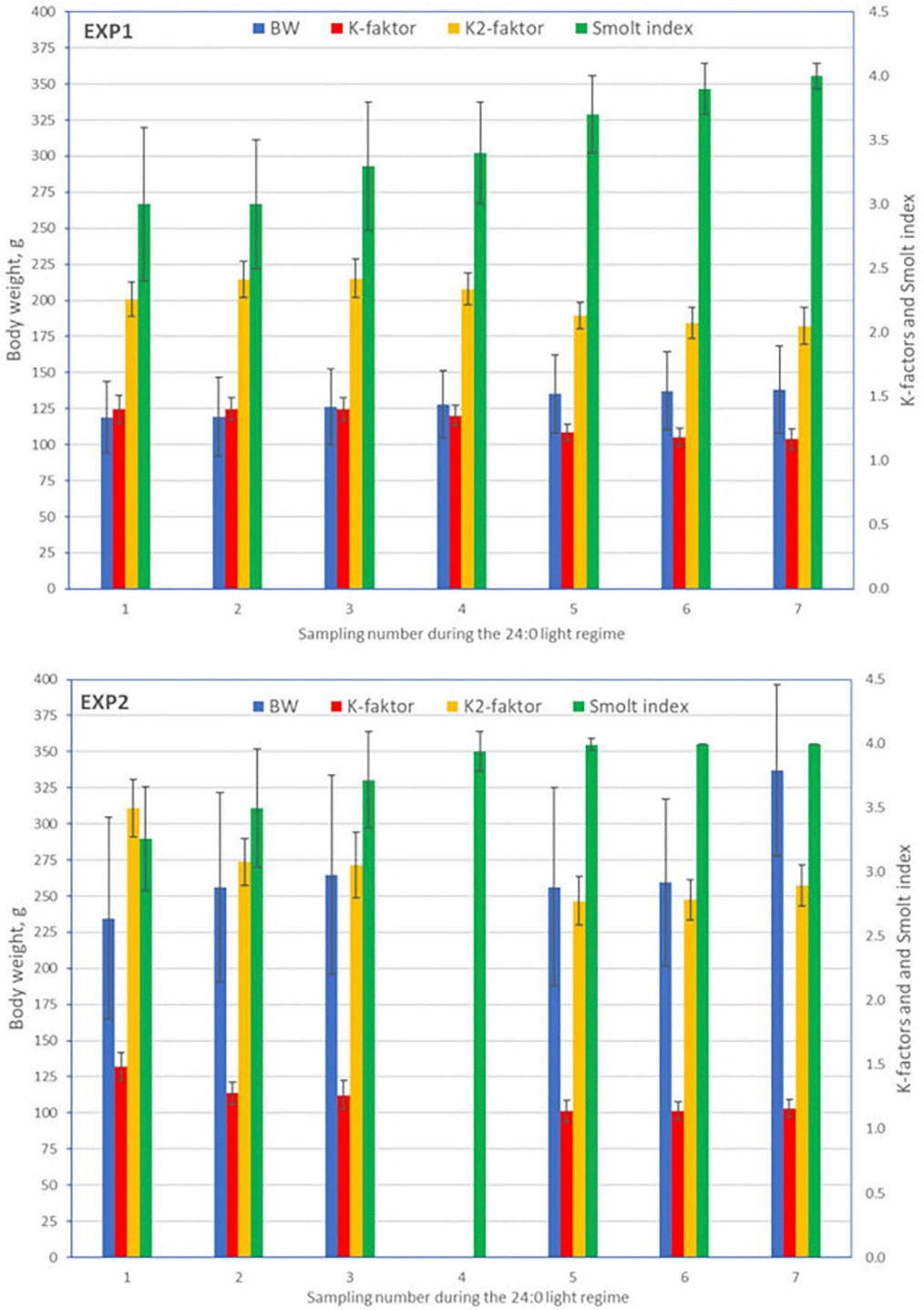

FIGURE 2 | Means (±standard deviations) for body weight, K-factor, K2-factor, and smolt index at each sampling for EXP1 and EXP2.

bivariate model similar to the model in 2.3.1. In the bivariate model, tank effect was fitted as the fixed effect for each of the traits. For the bivariate model with the two SI traits, full convergence was not achieved. The residual correlation between the same trait in the two experiments was set equal to zero as these traits were recorded on different animals.

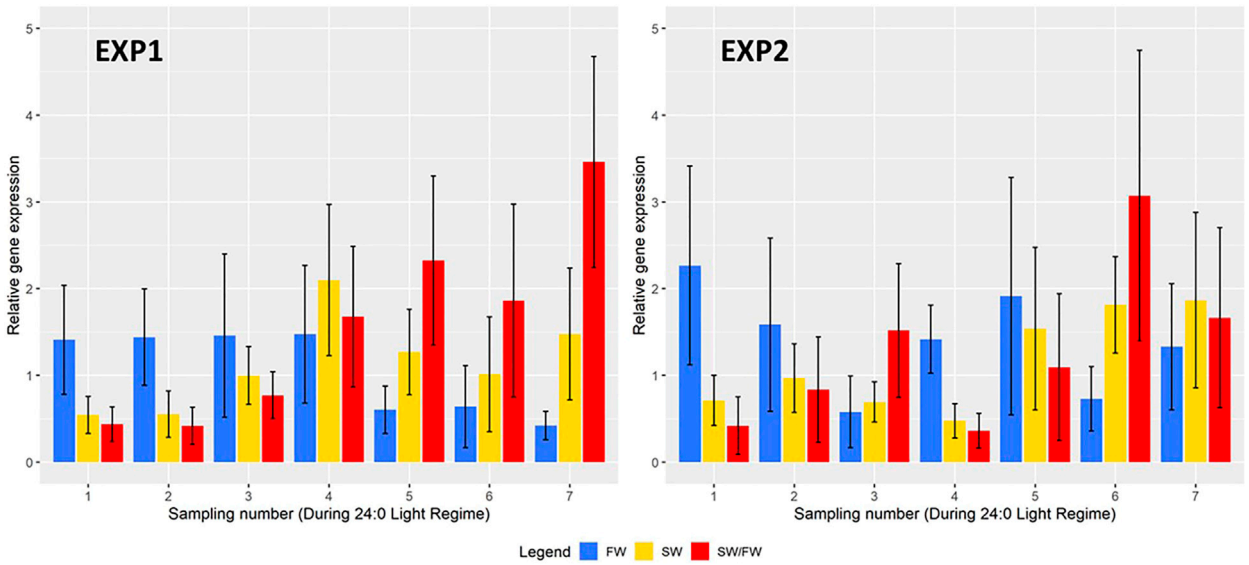

FIGURE 3 | Mean real-time quantitative polymerase chain reaction (RT-qPCR) values (±their standard deviations) of two key genes {Na^+K^+ATPase [NKA α1a for the freshwater variant (FW); NKA α1b for the saltwater variant (SW)]} with effect on Na^+/K^+ATPase in the gills and their ratio (SW/FW) at each sampling for EXP1 and EXP2.

RESULTS

Descriptive Statistics for the Traits Recorded at Each Sampling

A total of 1,595 (EXP1) and 1,773 (EXP2) fish with phenotypes were recorded from the seven samplings of each of the two experiments. Separately for EXP1 and EXP2, means and standard deviations at each sampling, for BW, K-factor, K2-factor, and SI are presented in **Figure 2**, and for FW, SW, and SW/FW in **Figure 3**.

Body Weight, K-Factor, K2-Factor, and Smolt Index

For EXP1, mean values of BW increased over the samplings, while for EXP2, the mean BW at both fifth and sixth sampling for unknown reason were lower than at the third sampling (BW at the fourth sampling had to be omitted due to some systematic measurement error of BW and BL seen as unreasonable K-factor values). K-factor decreased from about 1.4 to 1.2 (EXP1, 16.6%) and 1.5 to 1.2 (EXP2, 21.8%), and K2-factor from about 2.3 to 2.2 (EXP1, 9.2%) and 3.5 to 2.9 (EXP2, 17.1%). Due to unknown sampling error at the fourth sampling of EXP2, the mean values for BW, K-factor and K2-factor were removed from **Figure 2**. The Smolt index increased gradually from 3.0 (EXP1) and 3.2 (EXP2) until its maximum value of 4.0 at the seventh sampling of EXP1 and fifth sampling of EXP2 (**Figure 2**).

For BW, the coefficient of variation (CV) ranged from 17 to 30% (EXP1) and 18–23% (EXP2). K-factor had a small CV ranging from 5 to 8% for EXP1 and 6–9% for EXP2, while the CV for K2-factor was marginally lower. For SI, CV decreased from 20% (EXP1) and 12% (EXP2) at the first sampling to very low values (≤5%) at the sixth and seventh sampling (EXP1) and fourth, fifth, sixth, and seventh sampling (EXP2).

Real-Time Quantitative Polymerase Chain Reaction Values and Their Ratio

The relative gene expression of FW, SW, and SW/FW for each sampling is shown in **Figure 3** for EXP1 and EXP2. In EXP1, the mean FW was very similar (ranging from 1.41 to 1.48) at the first four samplings and decreased to a much lower value (ranging from 0.42 to 0.64) at the last three samplings. The mean SW increased gradually over the samplings and decreased to a lower value at the last three samplings. For SW/FW, the mean increased over the first five samplings, decreased to a lower value at the sixth sampling and increased to a higher mean value at the seventh sampling. At each of the seven samplings, these three traits showed a large variation among the animals as seen from their large standard deviations in **Figure 3** with CV ranging from 33 to 65%.

In EXP2, the mean values of all these three traits showed large variation from one sampling to the next, most probably due to a lesser number of fish recorded per sample (30–60 compared with 100 in EXP1) and the very large CV of the traits (28–79%). For the trait SW/FW, the mean value was higher at the last two samplings than at the fifth and major sampling.

The frequency distributions of FW, SW, and SW/FW are presented in **Figure 4**. For SW/FW, the variation among individuals increases over the smoltification period and were much larger during the last four compared with the three first samplings.

Correlations Between Traits Within Each Sampling

Figure 5 shows the correlations between some selected pair of traits within each of the seven samplings, separately for EXP1 and EXP2. The figure shows that the correlations vary over the

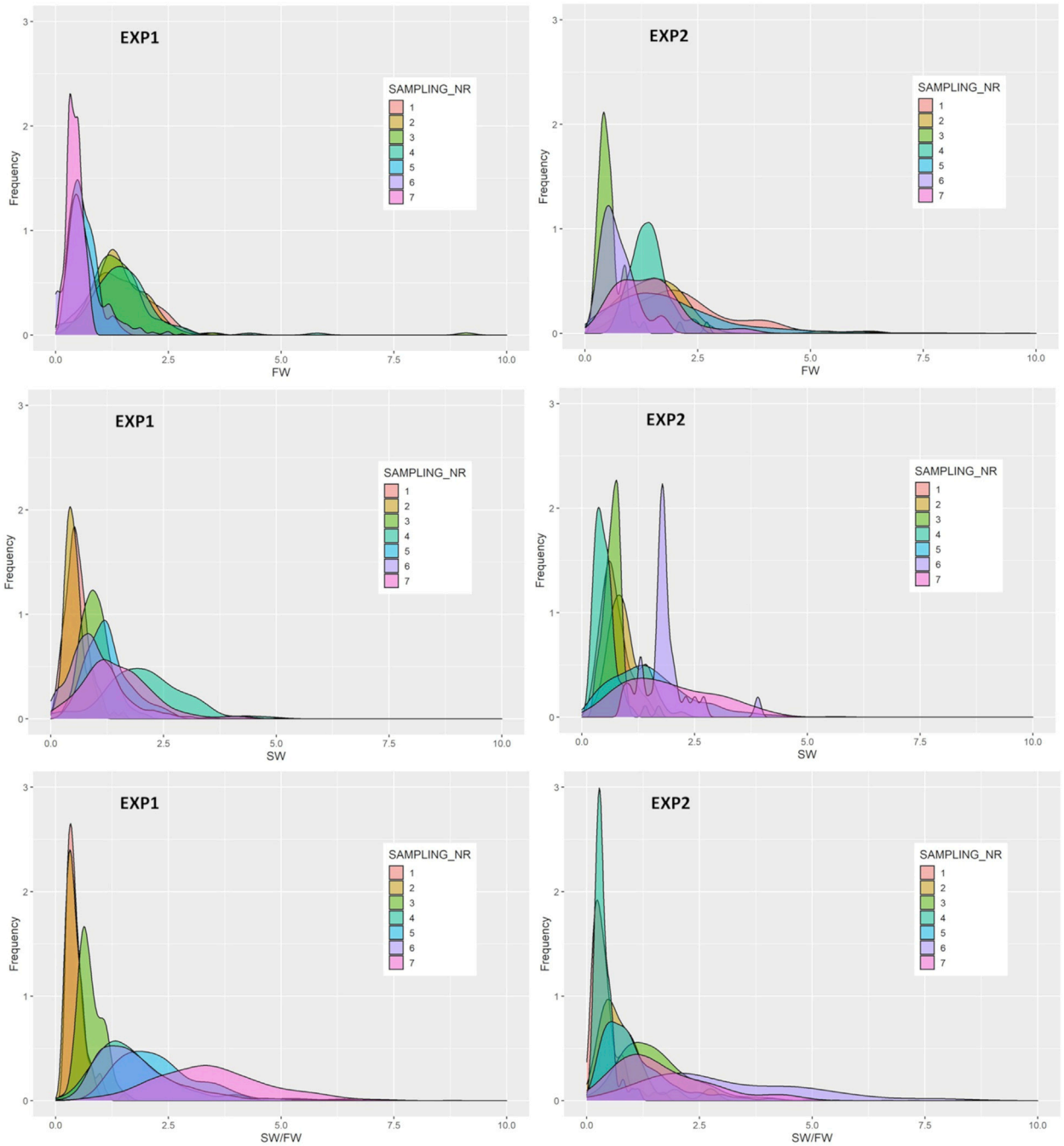

FIGURE 4 | Frequency distributions of the relative expression of freshwater (FW) and the seawater (SW) variants and their ratio (SW/FW) determined by two genes with effect on the Na^+K^+ATPase activity in the gills at each of the seven samplings of EXP1 and EXP2.

seven samplings and with different patterns for the two experiments.

Descriptive Statistics for the Traits at the Major Sampling

Descriptive statistics for the traits measured at the major samplings are presented in **Table 2**. Average weight at this sampling was 137.7 (SD = 27.6) and 256.4 g (SD = 68.6 g) for EXP1 and EXP2, respectively. Interestingly, SI was close to the maximum mean value of 3.95 (EXP1, sixth sampling) and 3.99 (EXP2, fifth sampling) showing that, morphologically, nearly all the fish were ready for seawater transfer at the major samplings. However, based on the mean SW/FW of 1.86 (EXP1) and 1.09 (EXP2), compared with the set desired benchmark of 2.00 (personal communication PHARMAQ Analytiq AS), the fish

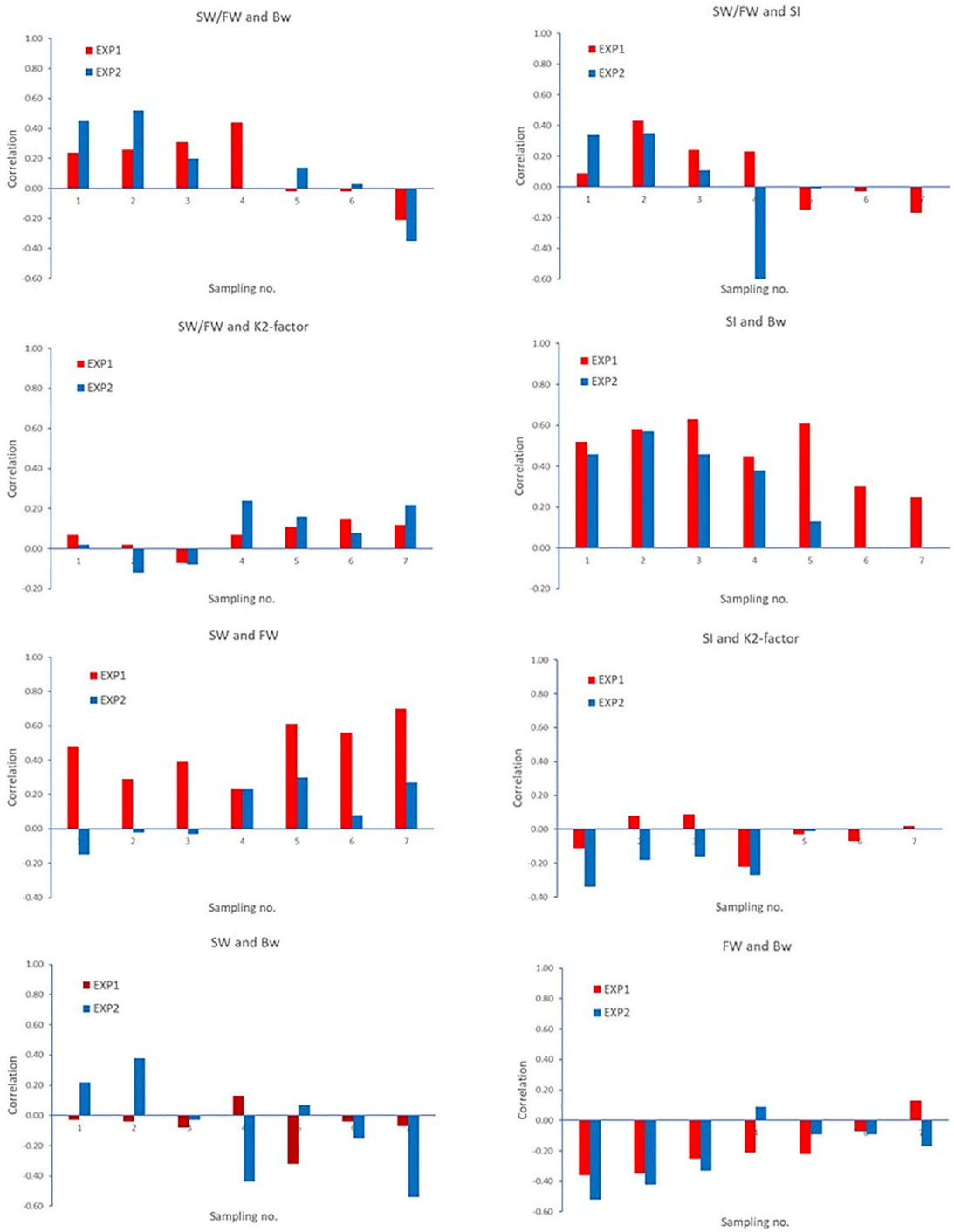

FIGURE 5 | Correlations between some selected pair of traits within each of the seven samplings. Correlation –0.254 > r > 0.254 (EXP1, N = 100 fish recorded/sampling) and –0.325 > r > 0.325 (EXP2, N = 2 × 30 fish recorded/sampling) are significantly different from zero ($p < 0.05$).

were physiologically (on average) near (EXP1) or far from (EXP2) being ready for seawater transfer. The very large CV of SW/FW (60% for EXP1 and 78% for EXP2) indicates substantial variation among the animals with respect to their readiness for seawater transfer.

Parameter Estimates for the Homologous Trait Between EXP1 and EXP2 at Major Sampling

In EXP1 and EXP2, similar bivariate heritability estimates were found for both BW and SW/FW, but for K2-factor, SI, FW, and SW, the heritability estimates differ by more than 50% between

TABLE 2 | Descriptive statistics for body weight (BW), body length (BL), condition factor (K-factor and K2-factor), freshwater variant (FW), seawater variant (SW), SW/FW ratio, and smolt index (SI) recorded at the major sampling of the fish in EXP1 (sampling 6) and EXP2 (sampling 5).

Trait	Experiment	N	Mean	Min	Max	SD	CVx100
BW	EXP1	944	137.7	11.6	215.7	27.6	20.1
	EXP2	15,00	256.4	91.1	556.0	68.6	26.8
BL	EXP1	944	22.6	13.5	26.4	1.59	7.0
	EXP2	1,500	28.4	19.8	37.0	2.62	9.2
K2-factor	EXP1	942	1.92	1.34	3.15	0.12	6.2
	EXP2	1,500	2.22	1.21	3.16	0.14	6.2
SI	EXP1	944	3.95	2.30	4.00	0.16	4.1
	EXP2	1,500	3.99	3.00	4.00	0.05	1.3
FW	EXP1	944	0.68	0.10	4.30	0.46	67.6
	EXP2	1,500	1.91	0.10	11.40	1.37	71.4
SW	EXP1	944	1.07	0.00	5.30	0.64	60.1
	EXP2	1,500	1.54	0.00	10.50	0.94	60.8
SW/FW	EXP1	944	1.86	0.00	12.00	1.11	59.6
	EXP2	1,500	1.09	0.00	7.25	0.85	77.5

the experiments (**Table 3**). For SI, the parameter estimates did not fully converge (and the estimates were therefore unreliable), as may be expected as nearly all fish were close to the maximum possible value of 4.0 for the trait in both EXP1 and EXP2 and, thus, with very low phenotypic variation (**Table 2**).

The genetic correlation between the homologous traits in EXP1 and EXP2 were high, i.e., 0.89 ± 0.04 for BW, 0.92 ± 0.06 for K2-factor, and 0.82 for SI (**Table 3**). However, for the other studied traits, the genetic correlations between the homologous traits in the two experiments were close to, and not significantly different from, zero.

Parameter Estimates for the Different Traits Within EXP1 and EXP2

The heritability estimates from the multitrait models were similar to those obtained from the bivariate models (results not shown). **Table 4** shows that the genetic correlations between BW and K2-factor was positive ranging from low to medium (0.22–0.50). Genetic correlations between BW and the smoltification traits (FW, SW, and SW/FW) were, in general, low and not significantly different from zero. However, the genetic correlation of K2-factor and the smoltification traits were low (not significantly different from zero, EXP2) to moderate (EXP1). Lastly, the genetic correlations between the two main smoltification traits (FW and SW) were low and not significantly different from zero, while the residual correlation between the two traits were positive.

TABLE 3 | Estimates of additive genetic and residual variances and heritability from bivariate models of the same trait in EXP1 and EXP2, and the genetic (r_g, upper line for each trait) and phenotypic (r_p, lower line for each trait) correlation between the same trait in the two experiments.

Trait	Experiment	σ^2_A ± s.e.	σ^2_e ± s.e.	h^2 ± s.e.	r_g/r_p
BW	EXP1	543 ± 67	335 ± 28	0.62 ± 0.04	0.89 ± 0.04
	EXP2	2,918 ± 314	1,806 ± 132	0.62 ± 0.04	0.55 ± 0.04
K2-factor	EXP1	0.01 ± 0.00	0.02 ± 0.00	0.27 ± 0.04	0.92 ± 0.06
	EXP2	0.02 ± 0.00	0.01 ± 0.00	0.69 ± 0.04	0.40 ± 0.06
SI[a]	EXP1	0.003	0.023	0.12	0.82
	EXP2	0.000	0.003	0.01	0.02
FW	EXP1	0.01 ± 0.01	0.20 ± 0.01	0.04 ± 0.03	0.22 ± 0.39
	EXP2	0.19 ± 0.06	1.61 ± 0.08	0.11 ± 0.03	0.22 ± 0.39
SW	EXP1	0.02 ± 0.01	0.39 ± 0.02	0.04 ± 0.03	0.09 ± 0.38
	EXP2	0.07 ± 0.03	0.79 ± 0.04	0.09 ± 0.03	–0.01 ± 0.02
SW/FW	EXP1	0.23 ± 0.07	1.03 ± 0.06	0.18 ± 0.05	–0.11 ± 0.23
	EXP2	0.09 ± 0.03	0.62 ± 0.03	0.12 ± 0.03	–0.02 ± 0.03

[a]*Estimates did not converge, and the estimates were, thus, not reliable.*

TABLE 4 | Estimates of residual (upper diagonal) and genetic (lower diagonal) correlations (±s.e.) between body measurements and smoltification status traits for EXP1 and EXP2 from a multivariate animal model.

Experiment	Trait	BW	K2-factor	FW	SW	SW/FW
EXP1	BW	—	0.26 ± 0.05	0.01	–0.01	–0.04 ± 0.05
	K2-factor	0.50 ± 0.09	—	0.04	–0.05	–0.06 ± 0.04
	FW	–0.49	–0.61	—	0.53	[a]
	SW	–0.25	0.66	–0.16	—	[a]
	SW/FW	0.07 ± 0.15	0.68 ± 0.14	[a]	[a]	—
EXP2	BW	—	0.32 ± 0.05	–0.13 ± 0.04	0.08 ± 0.04	0.14 ± 0.04
	K2-factor	0.22 ± 0.08	—	–0.13 ± 0.04	0.07 ± 0.04	0.16 ± 0.04
	FW	–0.04 ± 0.16	–0.24 ± 0.15	—	0.32 ± 0.03	[a]
	SW	0.14 ± 0.17	0.02 ± 0.16	0.03 ± 0.28	—	[a]
	SW/FW	0.16 ± 0.14	0.33 ± 0.12	[a]	[a]	—

[a]*For EXP1, the parameter estimates did not converge for the four-trait model with BW × K2-factor × FW × SW, and the estimates are, thus, not reliable.*

DISCUSSION

Overall Findings

The ratio of SW/FW showed low to moderate but significant genetic variation in both 0+ and 1+ smolts in Atlantic salmon at the major sampling, which indicates the possibility for selection for improved synchronization of smoltification status at the time where the largest proportion of the fish are considered to be smolt. However, the genetic correlation between SW/FW of 0+ and 1+ was not significantly different from zero indicating very little shared genetic variation in SW/FW in 0+ and 1+ fish.

Temporal Changes During Smoltification and Determining Seawater Transfer

Smoltification status is a temporal and reversable process, where fish can be smoltified and de-smoltified depending on the individual and the environmental conditions (Hoar, 1988; Nichols et al., 2008). In practice, controlled artificial lighting regimes are used to synchronize smoltification status in groups of fish so that the majority are ready for seawater transfer. However, we observed very different temporal progression of smoltification between the 0+ and 1+ fish in EXP1 and EXP2 as seen in **Figures 2** and **3** as well as the very different patterns in the phenotypic correlations between traits across time and EXP1 and EXP2 in **Figure 5**.

In EXP1, the smoltification of fish showed a trend of progressive increase in SI as well as a decrease in FW and concurrent increase in SW/FW. Based on the progression of SI and SW/FW, it was predicted at sampling point five that by sampling point six, the majority of the fish were ready for seawater transfer, and this is when the major sampling occurred. This progression can be seen in **Figure 5** where the phenotypic correlations between SW/FW and SI were moderate and positive until sampling five at which point they changed sign to moderate negative. Crucially, based on SW/FW at the major sampling of EXP1, 347 (with SW/FW ratio value of ≥2) of 944 fish (36.8%) would have been deemed ready for transfer to the sea, while based on SI, 924 (with a Smolt index value of ≥3.5) of 944 (97.9%) were deemed ready for transfer to the sea.

However, EXP2 showed differences between smoltification phenotypes through time and differences compared with EXP1. A steady increase in SI was observed in the first four samplings, which then remained stable until sampling seven. While FW steadily decreased, SW increased, and consequently, SW/FW also increased over the first three samplings, followed by an abrupt change at sampling four. This progress is very clear in **Figure 5** where the phenotypic correlations between SI and SW/FW were moderate and positive but suddenly changed sign to strongly negative at sampling point four. At this point in time, it was not clear if the changes at sampling four were due to random sampling error in the 60 fish or if the optimal time of seawater challenge might have been missed and seawater transfer (major sampling) should occur at point five. However, when all samplings in EXP2 are viewed, it appears that there was a progression toward smoltification in the first three samplings followed by asynchronous smoltification at sampling point four, although it is also possible there was de-smoltification at sampling four and then subsequent re-smoltification at sampling six and seven. These changes could also reflect stochastic differences due to sampling sizes of 60 fish in the seawater challenges for EXP2. However, the major sampling at point 5 of 1,500 fish is consistent with the temporal trend with the smaller sampling on 60 fish and, thus, supports the notion that we observed asynchronous smoltification in EXP2 and not differences due to sampling size. Importantly, only 602 (40.1%) out of 1,500 fish were deemed ready for seawater transfer at major sampling based on their individual SW/FW, while based on SI, almost all (99.7%) 1.500 (with a Smolt index value of ≥3.5) were ready for seawater transfer. Collectively, these findings suggest that there is discordance in predicting smoltification status by SI and SW/FW in both 0+ and 1+ smolt.

An important consideration is that the body size profile of 1+ Atlantic salmon smolt is not static. In the mid-1980s, 1+ smolt were typically 30–50 g at seawater transfer, and this increased to 70–120 g by the early 2000s (Bergheim et al., 2009) to 150–250 g at present (personal communication, Mowi ASA). Research report large losses of smolt 2–3 months post seawater transfer (Aunsmo et al., 2008) and that larger 0+ smolt appears to have better hypo-osmoregulatory ability (Handeland and Stefansson et al., 2001). In addition, evidence that the full switch to RAS (recirculating aquaculture system) production of smolts in the Faroe Islands after the year 2000 and the subsequent larger smolts resulted in a reduction in sea cage mortality rate (Joensen, 2008, Bergheim et al., 2009). This has led to considerable research interest in producing larger 450 to 1,000 g of 1+ smolts in land-based facilities, in attempts to improve survival and growth after seawater transfer (Ytrestøyl et al., 2020; Bergheim, 2012; Kolarevic et al., 2014). In contradiction, the present study found inconsistencies between different smoltification parameters in 0+ and the larger 1+ smolts during smoltification with evidence of possible asynchronous or de-smoltification in larger 1+ smolts. Suggesting that accurate prediction of smoltification status and synchronization of smoltification may be more challenging in larger smolts. Furthermore, SI and SW/FW gave contradictory indications of smoltification status in both 0+ and 1+ smolts, and given the wide array of smoltification regimes and smoltification phenotypes used in Atlantic salmon production, we speculate that contributing factors to the smolt mortality in the first 3 months after transfer to seawater may be due to incorrect determination of smoltification status and asynchronous smoltification status. This agrees with the findings of Kristensen et al. (2012) who surveyed mortality in the first 90 days post seawater transfer in commercial production in Norway and found increased mortality in 1+ compared with 0+ smolt despite a trend in increasing size of 1+ smolt. A more recent analysis of mortality in the Norwegian Atlantic salmon industry further confirmed higher mortalities in 1+ compared with 0+ smolts as well as an adverse effect of increasing body weight of smolts at seawater transfer (Oliveira et al., 2021).

Genetic Variation for Smoltification Status traits—Seawater Variant/Freshwater Variant and Smolt Index

The smolt index and SW/FW were the two primary traits used in this study for quantifying the smoltification status of the 0+ and 1+ fish subjected to a controlled artificial lighting regime. The differences between the 0+ and 1+ groups described above persisted at the genetic level. For example, SI had significant genetic variation for 0+ fish ($h^2 = 0.12$), whereas for the 1+ fish, SI had practically zero phenotypic and genetic variation. For the SW/FW, both 0+ and 1+ populations had moderate heritability, indicating significant additive genetic control on gill Na^+K^+ATPase activity, which could be used toward genetic improvement on readiness of smolts for seawater transfer. However, SW and FW had low nonsignificant heritability in EXP1 ($h^2 = 0.04$) but did have a low but significant heritability in EXP2 ($h^2 = 0.09–0.11$). To the best of our knowledge, there have been no similar studies on quantifying smoltification status by using SW/FW or SI on Atlantic salmon or any other salmonids. There are studies that categorized smoltification into a binary trait, which makes direct comparisons challenging but still provide evidence for genetic variation in smoltification status. The first of which was by Refstie et al (1977), who defined 1+ smolts as fish that exceed a particular size threshold and found this to be heritable (0.16 ± 0.05) in Atlantic salmon. The second study used a similar smoltification index to define a binary smoltification trait for captive 1+ Atlantic salmon smolts derived from wild fish two generations previously and found h^2 of 0.60 and 0.48 for fish reared on two temperature regimes of 2°C apart (Debes et al., 2020).

Interestingly, the SW/FW for 0+ and 1+ were not genetically correlated, and this has implications on how SW/FW could be incorporated in selective breeding schemes as it indicates these are genetically two distinct traits. Another consideration is that SW/FW ratio as it is used here is aimed at breeding for fish that shows better synchronization in smoltification status given their size and a standardized smoltification regime. As such, it may be that the genetic background to synchronizing smoltification could be different between 0+ and 1+ smolts given their size and age at the onset of a smoltification regime. The differences in the genetic background of SI and SW/FW traits between 0+ and 1+ could also be that SI does not reflect a change after the first smoltification and reflects a permanent phenotypic change on the time scale the fish were studied on, while the changes in expression in SW/FW could be more reactive and dynamic measurements, which can detect de-smoltification over short time spans. This trial made use of flow-through land-based systems with ambient water temperatures, which were different between EXP1 and EXP2 (**Figure 1**). It could be that different age groups may react differently to the environmental cues for smoltification, such as water temperatures and light regimes. For example, McCormick et al. (2000), found that the increase in gill Na^+K^+ATPase and decrease in K-factor was more advanced for fish reared in 10°C compared with those reared in 2°C water, when both were subjected to the same daylength. Similarly, Debes et al. (2020) report differences in the heritability estimates for a binary smoltification trait between fish cultured on water temperature regimes 2°C apart. This may explain the difference in terms of smolt window for 0+ and 1+ fish in our two experiments. More research on the genetic background of synchronized smoltification under different water temperatures, smolt sizes, and light regimes are certainly needed.

The Genetic Relationships Between Body Size and Smoltification Status

The earliest research into smoltification revealed a bimodal growth of Atlantic salmon parr with the larger parr smoltifying as 1+ and the smaller smoltifying as 2+ (Kristinsson et al., 1985). This phenomenon was linked to a body size threshold for smoltification (Elson., 1957; Wedemeyer et al., 1980). With developments in Atlantic salmon husbandry including improved nutrition and genetics, this body size was reached at an earlier age as 1+, and during the last 10–15 years also as 0+ smolt (Mørkøre and Rørvik, 2001). In the first generations of captive Atlantic salmon breeding, strong genetic relationships were observed between body weight and smoltification status (binary) (r_g 0.89–0.85; Refstie et al., 1977). In a recent study where F2 unselected offspring of wild-caught Atlantic salmon were reared under modern rearing conditions, an equally strong genetic correlation of 0.92 was found between body length and smoltification status of 1+ smolts (Debes et al., 2020). These findings suggest that selection for improved growth rate and, thus, increased size should result in improved smoltification. The study of Refstie et al (1977) and Debes et al. (2020) both reported smolt percentages in the range of 18–78%. However, based on SI, in the present study, smolt percentages were in the range of 98% for 0+ to 99.7% 1+. Despite finding large genetic variation for body size ($h^2 = 0.62$) in both 0+ and 1+ smolts, the estimated genetic correlations between body weight and smoltification status defined by SW/FW was weak (0.07 and 0.16) and not significantly different from zero. This suggests that the more recent and further selection for increased growth rate will not have improved overall smoltification rates, but instead, it has shortened the time needed for parr to reach the critical size threshold for smoltification. This suggests a genetic uncoupling of size and smoltification and would suggest that novel phenotypes more directly related to hypo-osmoregulation, like SW/FW, are needed to improve smolt quality.

In previous salmonid smoltification studies, K-factor is one of the traits commonly used as an indicator of smoltification status (Hoar, 1988; Sigholt et al., 1998; McCormick et al., 2000; Nichols et al., 2008). In general, fish that are ready for seawater transfer will have a reduced K-factor relative to contemporaries that are not ready, i.e., fish are growing relatively more in length than in weight from parr to smolt, toward a more streamlined body shape (Folmar and Dickhoff, 1980; Hoar, 1988; McCormick et al., 2000; Nichols et al., 2008). Selection for reduced K-factor as a means for improving the synchronization of smoltification status is not recommended as reduced K-factor goes against the primary breeding goal of Atlantic salmon for increased size and growth. However, genetic variation for the time of smoltification should result in significant heritability for both

K2-factor and SW/FW as found in this study. As K2-factor is expected to decrease (as found in both EXP1 and EXP2) and SW/FW is expected to increase (as found in EXP1, but variable in EXP2) toward the optimum time for sea transfer, the correlation between K2-factor and SW/FW is expected to be negative. However, in this study, the genetic correlation between these two traits was positive, and the residual correlation is close to zero. Furthermore, in both EXP1 and EXP2, the phenotypic correlations between these traits were, in general, low negative in the first three samplings but changed to positive in the last four samplings. These temporal changes in traits highlight one of the greatest challenges in smoltification research, i.e., lack of knowledge and understanding of the changes in phenotypes of individual fish during the smoltification transformation due to lack of data from repeated samplings of individuals. In addition, it highlights the importance of selecting the correct time point for phenotyping large numbers of fish needed for genetic evaluation studies, as sampling too early or too late can result in low phenotypic variation as seen for SI in EXP2. Furthermore, sampling at different times can have large implications for the genetic relationships between smoltification status and K2-factor.

Last, studies are needed to genetically link smoltification status traits like SW/FW with growth and survival of smolt at sea, particularly during the first weeks after seawater transfer, which are the ultimate breeding objectives of smoltification status. A limitation of the SW/FW phenotype is that it remains difficult to measure, as gill biopsy sampling must be conducted postmortem in some countries including Norway. This is due to limited information on the health and welfare of smolt post gill biopsy for phenotypes like SW/FW, upon which animal ethical licenses can be applied for and assessed. McCormik et al. (1993) demonstrated a noninvasive procedure for fill biopsy in Altantic salmon from 44 to 77 g with no resultant mortality and no significant effect on growth and salinity tolerance after 26 days. At present, the Norwegian Food Safety Authority Mattilsynet, (https://www.mattilsynet.no/) has granted animal ethic approval for two studies investigating growth and survival of post smolts at sea after gill biopsy for smoltification phenotypes. This avenue of future research may better define the use of smoltification phenotypes from live or postmortem gill biopsies.

CONCLUSIONS

The smoltification phenotype SW/FW showed low to moderate, but significant, genetic variation in both 0+ and 1+ smolts indicating that this could be used in selective breeding to improve synchronization of smoltification. However, the low genetic correlations between SW/FW in 0+ and 1+ fish indicate that these traits have a different genetic background depending on the size and age of the smolt. Furthermore, smoltification phenotypes showed temporal progression over the smoltification period, and this progression varied between 0+ and 1+ smolt highlighting the importance of correctly timing the point at which phenotypes are measured and cohorts are transferred to seawater. In addition, this also highlighted the need for further research into noninvasive methods of objectively measuring individual smoltification and subsequent survival and growth at sea.

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.

ETHICS STATEMENT

The animal study was reviewed and approved by Norwegian Food Safety Authority (Mattilsynet) with approval ID 13174 (EXP1) and 15310 (EXP2).

AUTHOR CONTRIBUTIONS

HLK analysed the data and drafted the article, GD, BG made substantial writing contributions. SB prepared the genotyping files. SB, BG, GD assisted in analysis and data curation. SB, EH, BG were involved in funding acquisition. SB, EH, BG were involved in the design of study, sampling, phenotype recording, and actively contributed in discussions. All authors read the article, gave suggestions and comments for the improvement, and approved the final article.

FUNDING

This study was made possible by the Research Council of Norway through project no. 267650 (SmoltFieldGenetics).

ACKNOWLEDGMENTS

The authors would like to thank Per Brunsvik for advice and experience in smoltification and experimental design. The authors would like to extend their gratitude to the staff at Nofima Research Station for Sustainable Aquaculture, in particular, Kjellrun Hoås Gannestad for the expertise and assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.696893/full#supplementary-material

REFERENCES

Aunsmo, A., Bruheim, T., Sandberg, M., Skjerve, E., Romstad, S., and Larssen, R. (2008). Methods for Investigating Patterns of Mortality and Quantifying Cause-specific Mortality in Sea-Farmed Atlantic salmon *Salmo salar*. *Dis. Aquat. Org.* 81, 99–107. doi:10.3354/dao01954

Bergheim, A., Drengstig, A., Ulgenes, Y., and Fivelstad, S. (2009). Production of Atlantic Salmon Smolts in Europe—Current Characteristics and Future Trends. *Aquacultural Engineering* 41 (2), 46–52.

Boulding, E. G., Culling, M., Glebe, B., Berg, P. R., Lien, S., and Moen, T. (2008). Conservation Genomics of Atlantic salmon: SNPs Associated with QTLs for Adaptive Traits in Parr from Four Trans-Atlantic Backcrosses. *Heredity* 101, 381–391. doi:10.1038/hdy.2008.67

Debes, P. V., Piavchenko, N., Erkinaro, J., and Primmer, C. R. (2020). Genetic Growth Potential, rather Than Phenotypic Size, Predicts Migration Phenotype in Atlantic salmon. *Proc. R. Soc. B.* 287, 20200867. doi:10.1098/rspb.2020.0867

Elson, P. F. (1957). The Importance of Size in the Change From Parr to Smolt in Atlantic Salmon. *Can. Fish Cult.* 21, 1–6.

Folmar, L. C., and Dickhoff, W. W. (1980). The Parr-Smolt Transformation (Smoltification) and Seawater Adaptation in Salmonids. *Aquaculture* 21, 1–37. doi:10.1016/0044-8486(80)90123-4

Handeland, S. O., and Stefansson, S. O. (2001). Photoperiod Control and Influence of Body Size on Off-Season Parr–Smolt Transformation and Post-Smolt Growth. *Aquaculture* 192 (2–4), 291–307.

Handeland, S. O., Imsland, A. K., Nilsen, T. O., Ebbesson, L. O. E., Hosfeld, C. D., Pedrosa, C., et al. (2014). Osmoregulation in Atlantic salmon *Salmo salar* Smolts Transferred to Seawater at Different Temperatures. *J. Fish. Biol.* 85 (4), 1163–1176. doi:10.1111/jfb.12481

Hoar, W. S. (1988). "4 the Physiology of Smolting Salmonids," in *Fish Physiology*. Editors W. S. Hoar and D. Randall (New York, USA: Academic Press), XIB, 275–343. doi:10.1016/s1546-5098(08)60216-2

Joensen, R. (2008). Resirkulering av vand i oppdrett [Recirculating aquaculture]. Presentation at Seminar of Recirculation of Water in Aquaculture, Sunndalsøra, Norway, February 27–28, 2008.

Kolarevic, J., Baeverfjord, G., Takle, H., Ytteborg, E., Britt, K. M. R., Reiten, B. K. M., et al. (2014). Performance and Welfare of Atlantic salmon Smolt Reared in Recirculating or Flow through Aquaculture Systems. *Aquaculture* 432, 15–25. doi:10.1016/j.aquaculture.2014.03.033

Kristinsson, J. B., Saunders, R. L., and Wiggs, A. J. (1985). Growth Dynamics during the Development of Bimodal Length-Frequency Distribution in Juvenile Atlantic salmon (*Salmo salar* L.). *Aquaculture* 45, 1–20. doi:10.1016/0044-8486(85)90254-6

Le Bras, Y., Dechamp, N., Krieg, F., Filangi, O., Guyomard, R., Boussaha, M., et al. (2011). Detection of QTL with Effects on Osmoregulation Capacities in the Rainbow trout (*Oncorhynchus mykiss*). *BMC Genet.* 12, 46. doi:10.1186/1471-2156-12-46

Le Cren, E. D. (1951). The Length-Weight Relationship and Seasonal Cycle in Gonad Weight and Condition in the Perch (*Perca fluviatilis*). *J. Anim. Ecol.* 16, 188–204. doi:10.2307/1540

McCormick, S. D. (1993). Methods for Nonlethal Gill Biopsy and Measurement of Na+, K+-ATPase Activity. *Can. J. Fish. Aquat. Sci.* 50 (3), 656–658. doi:10.1139/f93-075

McCormick, S. D., Moriyama, S., and Björnsson, B. T. (2000). Low Temperature Limits Photoperiod Control of Smolting in Atlantic salmon through Endocrine Mechanisms. *Am. J. Physiology-Regulatory, Integr. Comp. Physiol.* 278, R1352–R1361. doi:10.1152/ajpregu.2000.278.5.r1352

McCormick, S. D., Regish, A. M., Christensen, A. K., and Björnsson, B. T. (2013). Differential Regulation of Sodium-Potassium Pump Isoforms during Smolt Development and Seawater Exposure of Atlantic salmon. *J. Exp. Biol.* 216 (7), 1142–1151. doi:10.1242/jeb.080440

McCormick, S. D., Regish, A. M., and Christensen, A. K. (2009). Distinct Freshwater and Seawater Isoforms of Na+/K+-ATPase in Gill Chloride Cells of Atlantic salmon. *J. Exp. Biol.* 212 (24), 3994–4001. doi:10.1242/jeb.037275

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

Mørkøre, T., and Rørvik, K-A. (2001). Seasonal Variations in Growth, Feed Utilisation and Product Quality of Farmed Atlantic salmon (*Salmo salar*) Transferred to Seawater as 0+ Smolts or 1+ Smolts. *Aquaculture* 199 (1-2), 145–157. doi:10.1016/s0044-8486(01)00524-5

Nash, R. D. M., Valencia, H., and Geffen, a. J. (2006). The Origin of Fulton's Condition Factor—Setting the Record Straight. *Fisheries* 31 (5), 236–238.

Nichols, K. M., Edo, A. F., Wheeler, P. A., and Thorgaard, G. H. (2008). The Genetic Basis of Smoltification-Related Traits in *Oncorhynchus mykiss*. *Genetics* 179, 1559–1575. doi:10.1534/genetics.107.084251

Norman, J. D., Robinson, M., Glebe, B., Ferguson, M. M., and Danzmann, R. G. (2012). Genomic Arrangement of Salinity Tolerance QTLs in Salmonids: a Comparative Analysis of Atlantic salmon (Salmo salar) with Arctic Charr (Salvelinus alpinus) and Rainbow trout (Oncorhynchus mykiss). *BMC genomics* 13 (1), 420. doi:10.1186/1471-2164-13-420

Oliveira, V. H. S., Dean, K. R., Qviller, L., Kirkeby, C., and Bang Jensen, B. (2021). Factors Associated with Baseline Mortality in Norwegian Atlantic salmon Farming. *Sci. Rep.* 11, 1–15. doi:10.1038/s41598-021-93874-6

Piironen, J., Kiiskinen, P., Huuskonen, H., Heikura-Ovaskainen, M., and Vornanen, M. (2013). Comparison of Smoltification in Atlantic salmon (*Salmo salar*) from Anadromous and Landlocked Populations under Common Garden Conditions. *Ann. Zoologici Fennici* 50, 1–15. doi:10.5735/086.050.0101

Refstie, T., and Gjedrem, T. (1977). Selection Experiments With Salmon. II. Proportion of Atlantic Salmon Smoltifying at 1 Year of Age. *Aquaculture* 10.3 50, 231–242.

Sigholt, T., Åsgård, T., and Staurnes, M. (1998). Timing of Parr-Smolt Transformation in Atlantic salmon (*Salmo salar*): Effects of Changes in Temperature and Photoperiod. *Aquaculture* 160, 129–144. doi:10.1016/s0044-8486(97)00220-2

Stefansson, S. O., Björnsson, B. T., Ebbesson, L. O., and McCormick, S. D. (2008). "Smoltification," in *Fish Larval Physiology*. Editors R. N. Finn and B. G. Kapoor. (New Hampshire, United States of America: Science Publishers), 639–681.

Stefansson, S. O., McGinnity, P., Björnsson, B. T., Schreck, C. B., and McCormick, S. D. (2003). The Importance of Smolt Development to salmon Conservation, Culture, and Management: Perspectives from the 6th International Workshop on Salmonid Smoltification. *Aquaculture* 222, 1–14. doi:10.1016/s0044-8486(03)00098-x

Thorpe, J. E., Mangel, M., Metcalfe, N. B., and Huntingford, F. A. (1998). Modelling the Proximate Basis of Salmonid Life-History Variation, with Application to Atlantic salmon, *Salmo salar* L. *Evol. Ecol.* 12 (5), 581–599. doi:10.1023/a:1022351814644

Thorstad, E. B., Whoriskey, F., Uglem, I., Moore, A., Rikardsen, A. H., and Finstad, B. (2012). A Critical Life Stage of the Atlantic salmon *Salmo salar*: Behaviour and Survival during the Smolt and Initial post-smolt Migration. *J. fish Biol.* 81 (2), 500–542. doi:10.1111/j.1095-8649.2012.03370.x

Wedemeyer, G. A., Saunders, R. L., and Clarke, W. C. (1980). Environmental Factors Affecting Smoltification and Early marine Survival of Anadromous Salmonids. *Mar. Fish. Rev.* 42, 1–14.

Ytrestøyl, T., Takle, H., Kolarevic, J., Calabrese, S., Timmerhaus, G., Rosseland, B. O., et al. (2020). Performance and Welfare of Atlantic salmon, *Salmo salar* L. post-smolts in Recirculating Aquaculture Systems: Importance of Salinity and Water Velocity. *J. World Aquac. Soc.* 51 (2), 373–392.

Conflict of Interest: Author EH is employed by the company PHARMAQ Analytiq AS and SB is presently employed by the company Mowi ASA.

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 12 November 2021
doi: 10.3389/fgene.2021.778487

A High-Density Genetic Linkage Map and Fine Mapping of QTL For Feed Conversion Efficiency in Common Carp (*Cyprinus carpio*)

Xiaofeng Zhang, Peixian Luan, Dingchen Cao and Guo Hu**

National and Local United Engineering Laboratory for Freshwater Fish Breeding, Heilongjiang River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Harbin, China

OPEN ACCESS

Edited by:
Alexandre Wagner Silva Hilsdorf, University of Mogi das Cruzes, Brazil

Reviewed by:
Chun Zhang, Hunan Normal University, China
Changxu Tian, Guangdong Ocean University, China

***Correspondence:**
Xiaofeng Zhang
zhangxf.yu@163.com
Guo Hu
huguo@hrfri.ac.cn

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

Received: *17 September 2021*
Accepted: *22 October 2021*
Published: *12 November 2021*

Citation:
*Zhang X, Luan P, Cao D and Hu G (2021) A High-Density Genetic Linkage Map and Fine Mapping of QTL For Feed Conversion Efficiency in Common Carp (*Cyprinus carpio*). Front. Genet. 12:778487. doi: 10.3389/fgene.2021.778487*

Feed conversion efficiency (FCE) is an economically crucial trait in fish, however, little progress has been made in genetics and genomics for this trait because phenotypes of the trait are difficult to measure. In this study, we constructed a high-density and high-resolution genetic linkage map with 28,416 SNP markers for common carp (*Cyprinus carpio*) based on high throughput genotyping with the carp 250K single nucleotide polymorphism (SNP) array in a full-sib F_1 family of mirror carp (*Cyprinus carpio*) consisting of 141 progenies. The linkage map contained 11,983 distinct loci and spanned 3,590.09 cM with an average locus interval of 0.33 cM. A total of 17 QTL for the FCE trait were detected on four LGs (LG9, LG20, LG28, and LG32), explaining 8.9–15.9% of the phenotypic variations. One major cluster containing eight QTL (qFCE1-28, qFCE2-28, qFCE3-28, qFCE4-28, qFCE5-28, qFCE6-28, qFCE7-28, and qFCE8-28) was detected on LG28. Two clusters consisting of four QTL (qFCE1-32, qFCE2-32, qFCE3-32, and qFCE4-32) and three QTL (qFCE1-20, qFCE2-20, and qFCE3-20) were detected on LG32 and LG20, respectively. Nine candidate genes (*ACACA*, *SCAF4*, *SLC2A5*, *TNMD*, *PCDH1*, *FOXO*, AGO1, *FFAR3,* and *ARID1A*) underlying the feed efficiency trait were also identified, the biological functions of which may be involved in lipid metabolism, carbohydrate metabolism, energy deposition, fat accumulation, digestion, growth regulation, and cell proliferation and differentiation according to GO (Gene Ontology). As an important tool, high-density and high-resolution genetic linkage maps play a crucial role in the QTL fine mapping of economically important traits. Our novel findings provided new insights that elucidate the genetic basis and molecular mechanism of feed efficiency and the subsequent marker-assisted selection breeding in common carp.

Keywords: common carp (*Cyprinus carpio*), genetic linkage map, feed conversion efficiency, QTL mapping, candidate genes

INTRODUCTION

Aquatic products are an important source of nutrition for people around the world, especially in food-deficient countries (Easterling, 2007; Rice and Garcia, 2011). The long-term challenge for aquaculture breeders is to improve the yield-related traits in fish to meet the growing demand for fish products while minimizing their impact on the environment (Wringe et al., 2010; Laghari et al., 2013). As feed cost comprises about 65–75% of the total production cost in most aquaculture industries, an effective way to solve this problem is to breed fish with high feed conversion efficiency (FCE) (Gjedrem and Baranski, 2009). Generally, feed conversion efficiency is defined as the ratio of feed intake to weight gain in animals (Sherman et al., 2009; Ferreira et al., 2016). Similar to most yield-related traits, FCE is a heritable trait controlled by a series of genes, the environment, and their interactions, which has been confirmed in livestock, poultry, and fish (Herd et al., 2003; Cai et al., 2008; Begli et al., 2016; Lu et al., 2017; Fu et al., 2020; Miao et al., 2021). So, the aim of improvement in FCE could be achieved by breeding to select genetically superior animals.

Globally, as one of the most important fishes, common carp (*Cyprinus carpio*) is cultured in over 100 countries worldwide with over 4 million metric tons of global annual production (Bostock et al., 2010; http://www.fao.org/fishery/statistics/global-aquaculture-production/query/en). Because of the economic importance of common carp for the aquaculture industry, over the past decades, researchers have developed a variety of genomic resources and genetic tools to facilitate genetic improvement and breeding programs (Sun and Liang, 2004; Christoffels et al., 2006; Williams et al., 2008; Xu P. et al., 2014). Especially, the completion of genome sequencing of Songpu mirror carp and subsequent development of a 250,000 high-quality SNPs array chip which provides favorable tools for ultra-high density linkage map construction (Xu J. et al., 2014; Xu P. et al., 2014; Xu et al., 2019).

High-quality genetic linkage maps are essential tools for quantitative trait loci (QTL) mapping. In past decades, a number of linkage genetic map had been constructed and QTL mapping for many traits in the fish have been studied (Yue, 2014; Ashton et al., 2017). Common carp genetic researchers have also constructed a number of linkage maps in common carp based on different mapping families using SSR and SNP markers in recent two decades. These linkage maps have been widely used for QTL mapping of many economically important traits in common carp. Such as growth-related traits, disease resistance, meat quality, sex-determination traits, and so on have been successfully mapped (Jun Wang, 2012; Laghari et al., 2013; Kuang et al., 2015; Lv et al., 2016; Peng et al., 2016; Zheng et al., 2016; Jia et al., 2020; Su et al., 2020).

It is obvious that genetic improvements for the efficiency of feed utilization are important. So far, some DNA variants that play a role in the feed efficiency of poultry and livestock have been proposed by QTL mapping and association studies (Koning et al., 2003; Barendse et al., 2007; Sherman et al., 2009; Do et al., 2014; Mignon-Grasteau et al., 2015; Silva et al., 2019; Ye et al., 2020; Delpuech et al., 2021). However, the feed intake of each experimental individual is generally difficult to measure in fish. So, only a few QTL analyses on FCE traits have been conducted in fish and these QTL mapping studies may be insufficient because of the relatively low power of linkage analyses (Wang et al., 2012b; Lu et al., 2017; Barría et al., 2021). Recently, with the aid of the high throughput SNP genotyping array, an ultra-high density linkage map has revealed many important findings related to growth-related traits, sex dimorphism, and muscle quality traits in common carp (Peng et al., 2016; Zheng et al., 2016).

In the present study, using a 250,000 SNPs chip, a high-density linkage map with 28,416 SNP markers was constructed in a full-sib family of common carp, which is the highest density genetic linkage map for common carp so far. Using this map, we carried out the QTL fine mapping for the FCE trait. The candidate genes were also recognized from the genome regions of quatative trait loci. Furthermore, the results of our analysis wil provide a basis for genetically improving the feed efficiency of common carp in the future.

MATERIALS AND METHODS

Mapping Family and Phenotypic Measurements

Songpu mirror carp (SMC) is a strain derived from a European subspecies (*C. carpio*) of common carp, which is one of the most valuable fish species for freshwater breeding as well as one of the species that is highly promoted to culture in China. An F_1 full-sib family of SMC was produced at the Hulan Aquaculture Experimental Station of the Heilongjiang River Fisheries Research Institute, Harbin, China. A large number of Songpu mirror carp individuals ($n = 500$) were collected as brood fish, and their genetic distances were estimated using polymorphic microsatellite markers. Then a pair of female and male mature fish with suitable genetic distance were used as the maternal parent and the paternal parent to generate the experiment family (F_1) by artificial crossing. After hatching, approximately 3,000 offspring were raised in a pond under a standard feeding regime. A total of 150 progeny were randomly collected from the experimental population as the fish panel for feed conversion tests after 60 days post-hatching. The fish were stocked individually in a tank with a size of 90 cm × 85 cm × 70 cm in a series of re-circulating aquarium systems. All conditions in these tanks were regularly maintained throughout the experiment, i.e., water temperature was 22°C and water flow rate is 1 ms^{-1}. The juveniles were fed solely by complete carp extruded feed. Each experimental fish was equipped with a feed box. The pallet feeds used in this experiment contain 34% crude protein, 10% crude lipid, and 7% ash (**Supplementary Table S1**). In order to eliminate potential differences and as much as possible, we trained the experimental fish to adapt to the culture environment and reared them individually in the re-circulating aquarium tanks for 2 weeks before the feed conversion trial. During the feed conversion trail, the experimental family was fed three times (9:00, 13:00, and 17:00) a day. The feces in each tank were siphoned out daily and the water was changed

completely once a week. The residue of feed was siphoned out, recorded, and deducted from the feed weight supplied each day to acquire the accurate feed consumption of each fish.

Subsequent phenotypic measurements of feed conversion efficiency were made. Phenotypic data of the FCE were collected after 3 months of feeding trials. Briefly, we recorded body weight (BW) at the beginning (initial BW, BWI) and the end (final BW, BWF) of the feeding test. Total feed intake (FI) was recorded as the difference between the beginning and final weight of feed used during the test. Then, the FCE was calculated as the BW gain after the experiment divided by total FI. Since the phenotypically extreme or abnormal individuals were excluded from the experiment trial according to the statistical analysis, a total of 141 offspring was determined as the experimental sample for linkage map construction and QTL analysis.

DNA Extraction and SNP Genotyping

Approximate 0.5 ml of blood from each sample was collected into a tube containing EDTA. Genomic DNA was extracted from the preserved blood using QIAamp DNA BloodMini Kit (Qiagen, Shanghai, China) following the manufacturer's protocol. A NanoDrop-1000 spectrophotometer (Thermo Fisher Scientific, United States) was used to determine the DNA concentration in each sample and the integrity of DNA was checked by 1.5% agarose gel electrophoresis. DNA samples used for genotyping were diluted to 50 ng/ul and genotyped at GeneSeek (Lincoln, Nebraska, United States) using the common carp 250K SNP array (Xu J. et al., 2014). Affymetrix CEL files were analyzed using Affymetrix Genotyping Console software (version 4.0) for quality control and genotype calling. The CHP files generated from Affymetrix Genotyping Console were then extracted and converted to Ped/Map format for further analysis. SNPs were removed if they had a missing genotype rate >1% and a minor allele frequency (MAF) < 1%. SNPs retained were collected for subsequent analysis.

Linkage Map Construction

Prior to map construction, Mendelian inheritance errors also were checked by a chi-square test using the parameters of segregation distortion ($p < 0.001$). Only the SNPs conforming to Mendelian inheritance were used for further linkage analysis. Then, the remaining markers through a series of quality control procedures above were subjected to JoinMap for linkage map construction. The double pseudo-test cross strategy was employed for linkage analysis. SNPs were separated into three segregation patterns: AAxAB or BBxAB (1:1 segregation only in male parent), ABxAA or ABxBB (1:1 segregation only in female parent), and ABxAB (1:2:1 segregation in both parents). The linkage maps were constructed by using JoinMap 4.0 (Van Ooijen, 2006) with "CP" type population, which is designed to handle F_1 population data containing various genotype configurations. A threshold of 5.0 was set for assigning markers into different linkage groups (LGs). The Kosambi mapping function was used to estimate map distances in centiMorgans (cM) through the maximum likelihood (ML) algorithm. Graphical visualization of the linkage maps was drawn by MapChart 2.2 software (Voorrips, 2002).

TABLE 1 | Descriptive statistics of phenotypic data.

Trait	Minimum	Maximum	Mean ± SD	CV
BWI (g)	15.65	42.85	26.52 ± 4.66	0.18
BWF (g)	42.99	271.52	150.72 ± 44.06	0.29
FCE (%)	40.8	89.4	60.8 ± 10.6	0.17
FI (g)	67.1	324.3	200.1 ± 41.3	0.21

FCE, feed conversion efficiency; BWI/BWF, initial/final body weight; FI, feed intake.

QTL Mapping and Annotation of Candidate Genes

QTL mapping analysis was performed for the FCE trait using software package MapQTL5.0 (Van Ooijen, 2004) with CIM (composite interval mapping) and MQM (multiple QTL model) mapping algorithms. A 1,000 permutation test was used to determine the LOD score significance thresholds at a 95% confidence level. After the 1,000 permutation test, a LOD threshold of 2.8–3.1 was set to identify significant QTL on each linkage group. The phenotypic variance explained (PVE) was estimated through stepwise regression (Li et al., 2007).

For each nearest SNP marker of QTL, we extracted candidate genes at the SNP loci from the reference genome of common carp. To annotate the functions of the FCE genes, we searched their orthologs by blastx against eudicots non-redundant database with an e-value threshold of 10^{-5}. We also used Blast2GO (Conesa et al., 2005) with default parameters to assign the Gene Ontology (GO) to obtain more information of candidate genes that may be related to feed efficiency based on the annotation information.

RESULTS

Phenotypic Data

Out of 150 experimental fish fed in individual tanks, 141 were alive throughout the 3 months of the experiment and used for the further phenotyped analysis of feed efficiency. The descriptive statistics of the phenotypic measurements of feed efficiency used for the present studies are given in **Table 1**. The panel had 26.52 ± 4.66 g mean initial body weight (BWI). The minimum and maximum for the BWI were 15.65 and 42.85 g, respectively. After the 3-month feeding trial, the average final body weight (BWF) of the individuals reached 150.72 ± 44.06 g. The minimum and maximum for the BWF were 42.99 and 271.52 g, respectively. The total feed intake (FI) of the fish was between 67.1 and 324.3 g with an average value of 200.1 ± 41.3 g. The deduced feed conversion efficiency (FCE) of the 141 individuals ranged from 40.8 to 89.4% with an average value of 60.8% (SD = 10.6%). The coefficient of variation (CV) of four traits ranged from 0.17 to 0.29 (**Table 1**).

Selection of SNP Markers

The genotypic data of all 141 F_1 offspring and their parents is available. A total of 219,902 SNP markers were successfully genotyped. According to the assessment of genotyping quality and polymorphism in all samples from the mapping family, the

TABLE 2 | SNPs selected for linkage mapping.

Item	Number
SNPs on array	24,9913
SNPs successful genotyped	219,902
Polymorphic SNPs	102,741
SNPs with high genotyping quality	35,505
SNPs used for linkage mapping after further filtering	28,831
SNPs mapped to linkage map	28,416

genotypes of a total of 102,741 SNPs were exhibited polymorphic among the mapping panel. After further filtration with more stringent conditions to remove SNPs with low calling rate [SNPs calling rate 99% and minor allele frequency (MAF) greater than 1%], a total of 35,505 SNP markers were retained for further analysis. We selected 28,831 SNP markers based on segregation distortion and non-Mendelian inheritance ($p < 0.001$) for further linkage analysis and mapping (**Table 2**).

Linkage Map Construction

Among 28,831 SNP markers, 28,416 SNPs were mapped on 11,983 distinct positions in 50 linkage groups (**Figure 1**; **Table 3**; **Supplementary Table S2**). A total of 98.6% of high-quality SNPs markers could be successfully mapped. The remaining 415 markers were not mapped. The total map distance was 3,590.09 cM with an average value of 71.80 cM. The number of markers on the linkage group varied from 213 (LG23) to 920 (LG45), and the average number of mapped markers per LG was 568 markers. The genetic length of each LG ranged from 53.48 cM (LG37) to 159.93 cM (LG7) with an average length of 71.80 cM. The average locus intervals varied from 0.14 cM in LG31 to 0.58 cM in LG7. The overall average marker interval was 0.33 cM. Based on the method described by Chakravarti et al. (1991), the expected genome length was estimated to be 3,623.71 cM. So, the percentage of the genome covered by the linkage map was calculated to be 99.07%. Detailed information and characteristics of this high-density genetic map were summarized in **Table 3**. As shown in **Figure 2**, the SNP distribution on each linkage group was also examined, which illustrated an even distribution of SNP markers on each linkage group with some exceptions at the middle and terminal regions of LGs.

QTL Mapping

The profiles and characteristics of the QTL associated with FCE are presented in **Table 4**; **Figures 3, 4**. A total of 17 significant QTL regions were mapped onto four LGs using composite

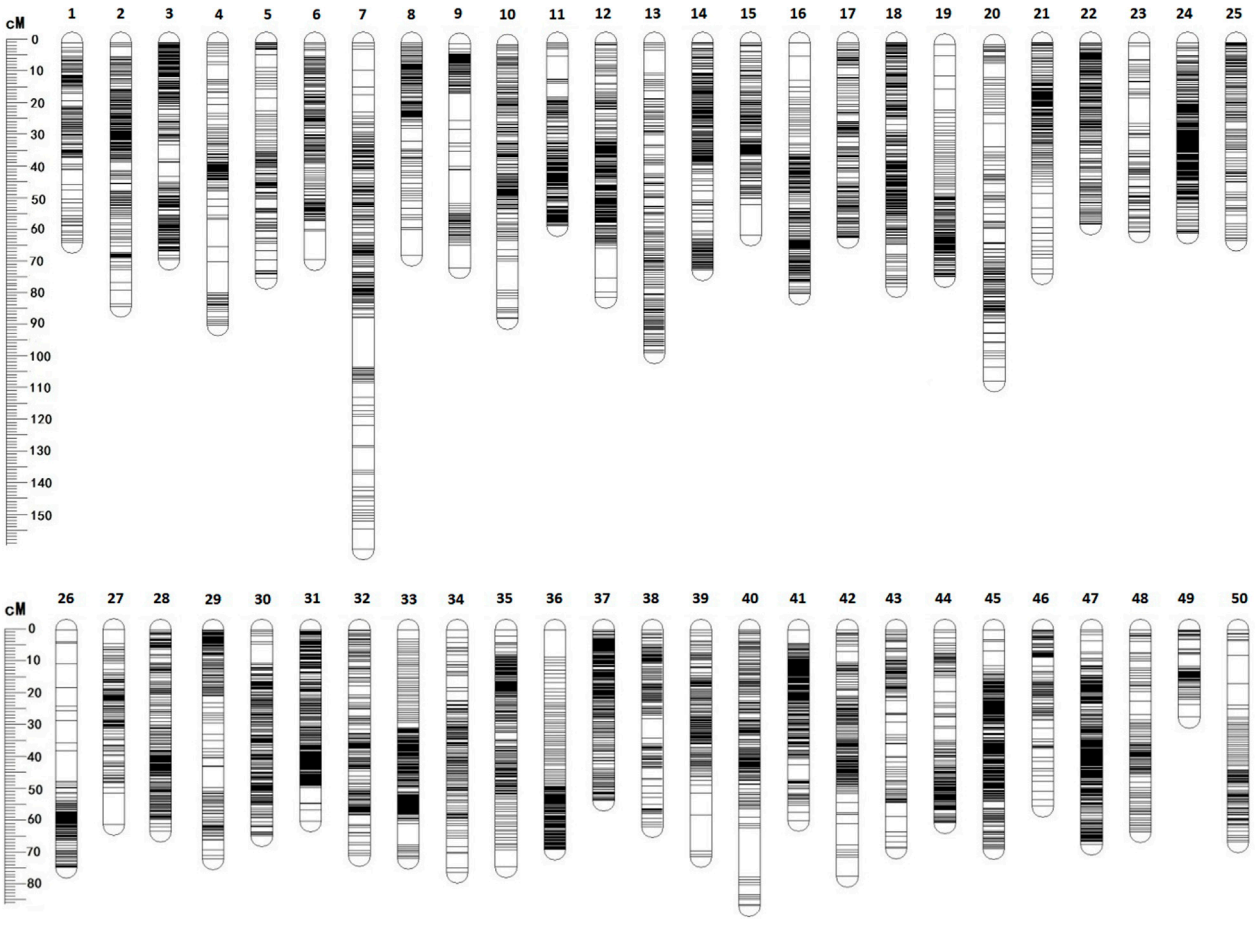

FIGURE 1 | The high-density linkage genetic map for mirror carp.

TABLE 3 | Summary of the linkage map of the mirror carp.

LG	No. of SNPs	Distinct positions	Genetic length (cM)	Locus interval (cM)
1	444	153	63.14	0.41
2	726	279	83.48	0.30
3	703	296	68.48	0.23
4	489	157	89.21	0.57
5	417	154	74.39	0.48
6	647	206	68.51	0.33
7	730	276	159.93	0.58
8	439	164	67.22	0.41
9	291	137	70.76	0.52
10	691	274	86.61	0.32
11	817	308	57.75	0.19
12	791	337	80.55	0.24
13	382	207	98.09	0.47
14	758	305	71.76	0.24
15	467	186	60.74	0.33
16	582	254	79.46	0.31
17	457	200	61.54	0.31
18	873	414	77.04	0.19
19	437	199	73.42	0.37
20	507	198	106.16	0.54
21	735	250	73.06	0.29
22	865	308	57.33	0.19
23	213	108	59.75	0.55
24	847	388	60.14	0.16
25	645	235	62.38	0.27
26	430	189	74.39	0.39
27	361	150	61.11	0.41
28	583	260	63.05	0.24
29	521	198	71.75	0.36
30	714	286	64.56	0.23
31	858	438	59.99	0.14
32	525	242	70.72	0.29
33	862	386	71.56	0.19
34	288	160	76.03	0.48
35	576	302	74.21	0.25
36	572	258	68.84	0.27
37	691	333	53.48	0.16
38	436	196	61.66	0.31
39	403	193	71.10	0.37
40	572	282	68.58	0.24
41	745	323	59.80	0.19
42	722	254	77.35	0.30
43	450	177	68.35	0.39
44	515	209	60.55	0.29
45	920	391	68.53	0.18
46	302	138	55.27	0.40
47	384	227	72.26	0.32
48	314	126	63.22	0.50
49	378	136	76.37	0.56
50	341	136	66.49	0.49
Total	28,416	11,983	3,590.09	0.33

interval mapping (CIM) and multiple QTL models (MQM) in MapQTL 5.0 program. The LOD significance thresholds for the FCE ranged from 2.8 to 3.1 based on the permutation test results. The phenotypic variances explained by each QTL (R_2) were estimated by CIM. As shown in **Table 3**, these regions include 51 SNP loci, distributed on four LGs including LG9, LG20, LG28, and LG32 and having an effect of 8.9–15.9% phenotypic variance explained (PVE). Most of these QTL were clustered together on their respective LGs. One major cluster containing eight QTL (qFCE1-28, qFCE2-28, qFCE3-28, qFCE4-28, qFCE5-28, qFCE6-28, qFCE7-28, and qFCE8-28) was detected between the narrow position of 37.07–44.78 cM on LG28, accounting for 9.5–10.6% PVE. Among them, qFCE8-28 located at 44.52–44.78 cM had the highest LOD value (3.44) and correspondingly had the highest contribution to phenotypic variation (10.6%). On LG32, a cluster situated within a region (15.07–32.64 cM) consisted of four QTL (qFCE1-32, qFCE2-32, qFCE3-32, and qFCE4-32) with a LOD value of 3.04–3.49 and was able to explain 9.5–12.1% of the PVE. The most significant QTL qFCE2-32 located on LG32 at 16.29–20.45 cM presented the highest LOD value of 3.49,

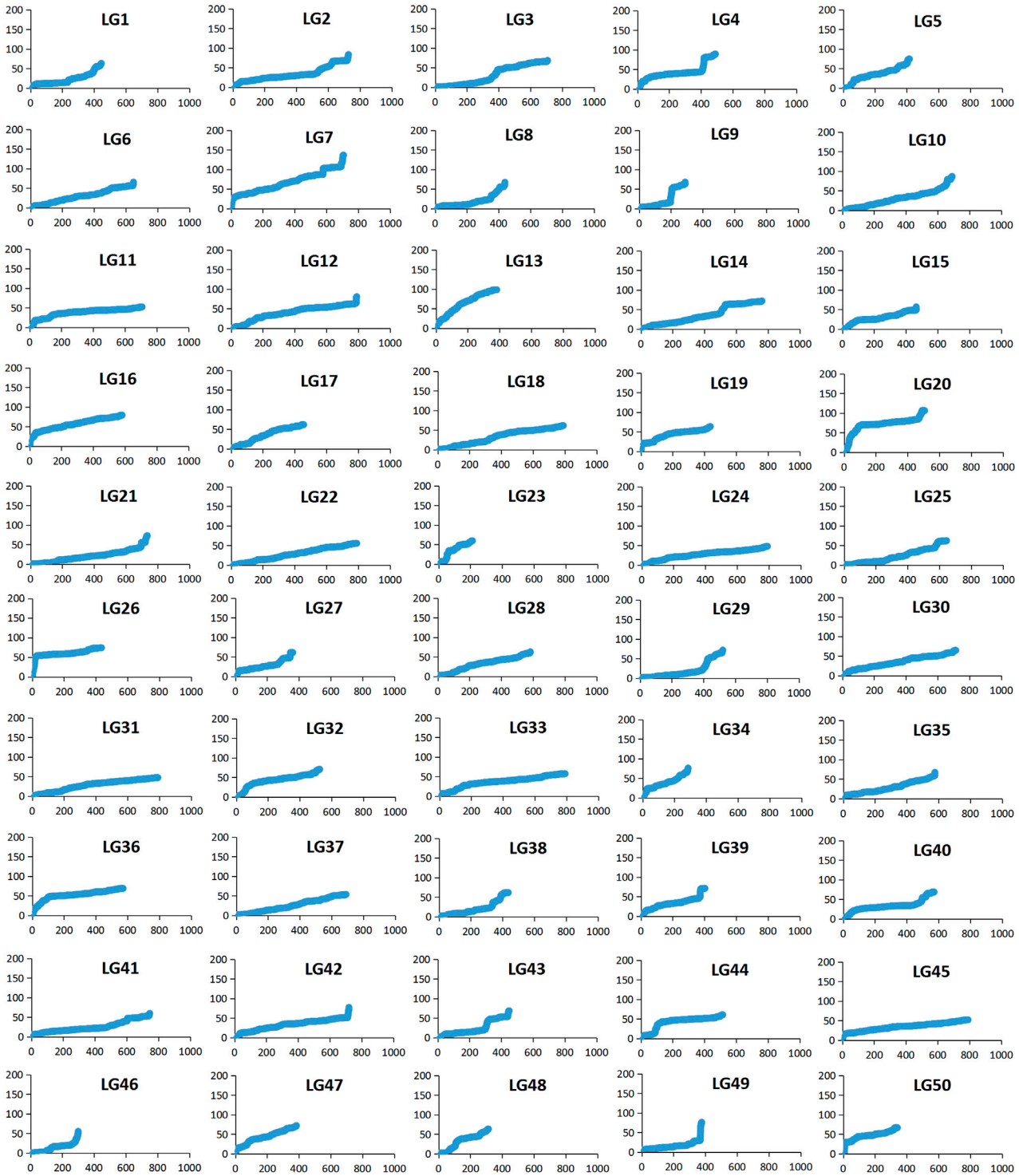

FIGURE 2 | The patterns of marker distribution on each linkage group. The X-axis represents marker orders on each linkage group. The Y-axis represents SNP marker position (cM) or each linkage group.

explaining 12.1% of the total PVE. Three QTL on LG20 (qFCE1-20, qFCE2-20, and qFCE3-20) consisted of a cluster located on 39.41–51.37 cM regions with LOD values of 2.85–2.96, and contributions to PVE of 8.9–9.2%. Two QTL (qFCE1-9 and qFCE2-9) on LG9 were detected at positions 15.81–27.08 and 32.32–38.81 cM, with LOD values of 3.20 and 3.11. The QTL qFCE2-9 located at LG9 at 32.32–38.81 cM explained the highest percentage of the total PVE of 15.9%.

TABLE 4 | Analysis of QTL and estimation of genetic effects.

Linkage group	QTL name	Marker interval	Marker interval (CM)	Linkage group size (CM)	LOD	Permutation*	PVE %	Nearest maker
LG9	qFCE1-9	snp108444-snp022617	15.81–27.08	70.76	3.20	3.1	12.9	snp204986
LG9	qFCE2-9	snp105119-snp098735	32.32–38.81	70.76	3.11	3.1	15.9	snp057963
LG20	qFCE1-20	snp049074-snp067360	39.41–41.65	106.16	2.91	2.8	9.1	snp221731
LG20	qFCE2-20	snp041990-snp145914	41.81–42.35	106.16	2.85	2.8	8.9	snp145912
LG20	qFCE3-20	snp145904-snp146285	49.45–51.37	106.16	2.96	2.8	9.2	snp145909
LG28	qFCE1-28	snp247225-snp164238	37.07–37.90	63.05	3.38	3.0	10.4	snp247112
LG28	qFCE2-28	snp060450-snp164244	38.03–38.19	63.05	3.26	3.0	10.1	snp060453
LG28	qFCE3-28	snp050416-snp173574	38.56–39.20	63.05	3.12	3.0	9.7	snp247094
LG28	qFCE4-28	snp094907-snp190172	39.67–40.04	63.05	3.13	3.0	9.7	snp094899
LG28	qFCE5-28	snp105167-snp204381	42.27–42.67	63.05	3.05	3.0	9.7	snp137852
LG28	qFCE6-28	snp025842-snp168930	43.09–43.28	63.05	3.28	3.0	10.2	snp164078
LG28	qFCE7-28	snp031019-snp123627	43.46–43.86	63.05	3.05	3.0	9.5	snp173520
LG28	qFCE8-28	snp006951-snp008550	44.52–44.78	63.05	3.44	3.0	10.6	snp171092
LG32	qFCE1-32	snp076503-snp004530	15.07–16.29	70.72	3.40	3.0	11.0	snp004531
LG32	qFCE2-32	snp004530-snp066978	16.29–20.45	70.72	3.49	3.0	12.1	snp004486
LG32	qFCE3-32	snp066980-snp157023	21.35–23.59	70.72	3.04	3.0	9.5	snp157021
LG32	qFCE4-32	snp189518-snp085154	31.09–32.64	70.72	3.13	3.0	9.8	snp193531

**Represents the chromosome-wide significance LOD, threshold at* $p < 0.05$.

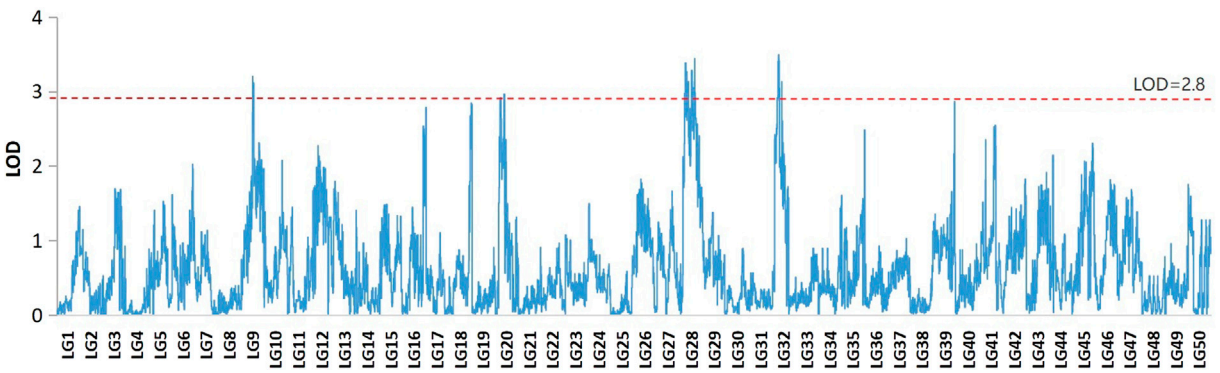

FIGURE 3 | QTL mapping and significant regions were identified for FCE in common carp.

Candidate Gene Identification

To further identify potential causative genes, we screened the genome and collected protein-coding genes at the nearest SNP loci from each QTL. A total of nine genes associated with feed conversion efficiency in nine QTL were identified (**Table 5**). These candidate protein-coding genes were annotated by GO. We identified a gene, acetyl-CoA carboxylase alpha (*ACACA*), from the qFCE1-9 on LG9. We also identified two genes, SR-related

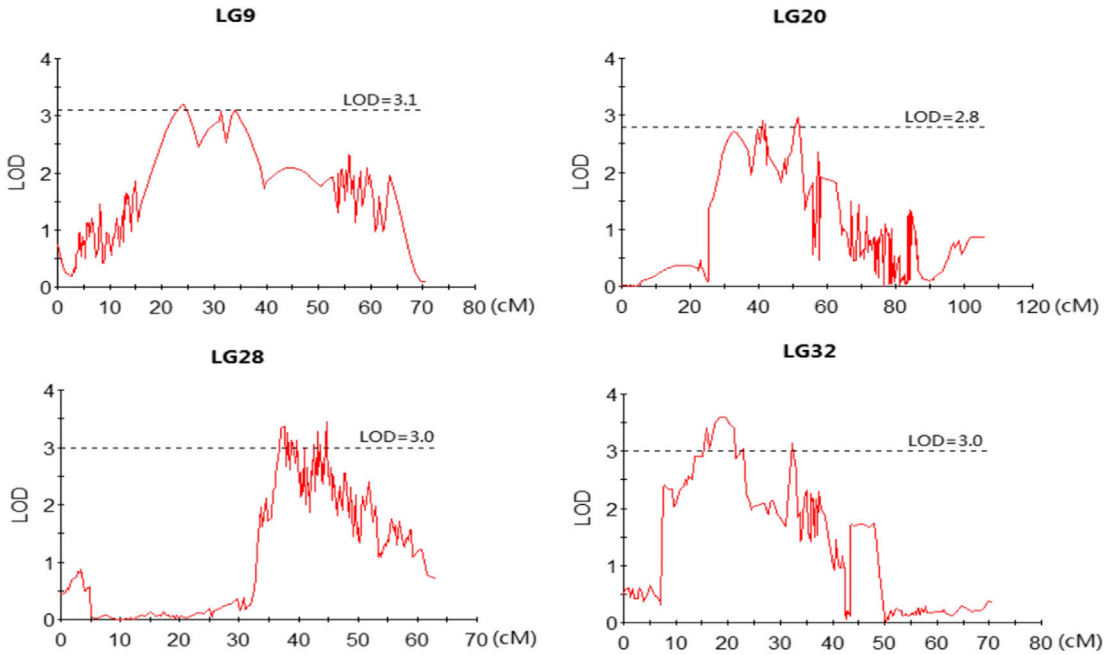

FIGURE 4 | LOD curves for QTLs contributing to FCE. The X-axis indicates marker distance and the Y-axis represents the LOD with the dashed line indicating the threshold value of permutation.

TABLE 5 | Summary of candidate genes for FCE trait in mirror carp.

QTL name	Chr	Associated SNPs	SNP location (bp)	Carp gene Id	Gene location (bp)		Gene name	Annotation
					From	To		
qFCE1-9	LG9	snp108444	22,188,096	CAFS_CommonC_G_091,796	22,185,301	22,207,580	*ACACA*	Acetyl-CoA carboxylase alpha
qFCE2-20	LG20	snp145912	13,067,352	CAFS_CommonC_G_057,517	13,064,285	13,078,405	*SCAF4*	SR-related CTD-associated factor 4
qFCE3-20	LG20	snp145909	14,807,639	CAFS_CommonC_G_057,228	14,795,909	14,807,894	*SLC2A5*	Solute carrier family 2 member 5
qFCE1-28	LG28	snp247112	10,256,532	CAFS_CommonC_G_063,176	10,254,148	10,297,988	*TNMD*	Tenomodulin
qFCE7-28	LG28	snp173520	10,000,245	CAFS_CommonC_G_062,723	9,998,381	10,015,209	*PCDH1*	Protocadherin 1
qFCE8-28	LG28	snp171092	2,535,708	CAFS_CommonC_G_062,559	2,533,722	2,542,477	*FOXO*	Forkhead box O
qFCE1-32	290	snp004531	562,238	CAFS_CommonC_G_004,026	560,784	589,106	*AGO1*	Argonaute RISC component 1
qFCE2-32	290	snp004486	372,623	CAFS_CommonC_G_004,020	370,723	376,820	*FFAR3*	Free fatty acid receptor 3-like
qFCE3-32	LG32	snp157021	20,092,699	CAFS_CommonC_G_068,410	20,081,844	20,104,892	*ARID1A*	AT rich interactive domain 1A

CTD-associated factor 4 (*SCAF4*) and solute carrier family 2 member 5 (*SLC2A5*), from the qFCE2-20 and qFCE3-20 on LG20, respectively. On LG28, we identified three genes at three QTL. At the qFCE1-28, we identified a tenomodulin (*TNMD*) gene of which it has been reported that its polymorphisms were associated with adiposity and also with glucose metabolism in men (Tolppanen et al., 2007). We further identified a gene, protocadherin 1 (*PCDH1*), at the qFCE3-28. We also recognized a forkhead box O (*FOXO*) gene from the qFCE8-28. This gene encodes a protein that is regulated by factors involved in growth and differentiation indicating it plays a role in these processes (Wang et al., 2009; Zhu et al., 2010; Huo et al., 2014; Watamoto et al., 2019). We also identified three genes, argonaute RISC component 1 (AGO1), free fatty acid receptor 3-like (*FFAR3*), and AT rich interactive domain 1A (*ARID1A*) at qFCE1-32, QTL qFCE2-32, and qFCE3-32 on LG32, respectively.

DISCUSION

In past decades, fish-breeding methodologies have developed from traditional selection to modern biotechnologies, such as marker-assisted selection (MAS) and molecular breeding (Yue, 2014; Tong and Sun, 2015; Ashton et al., 2017). Implementation of MAS requires DNA markers that are tightly linked to traits of interest by means of QTL mapping or association analysis (Lande and Thompson, 1990; Rezende

et al., 2012). Most economically important traits in fish, such as growth, disease resistance, and sex are controlled by multiple genes known as QTL. Most of these QTL have minor effects, but several pyramided may have major effects on traits (Tong and Sun, 2015). Theoretically, if genes and genetic markers associated with traits of interest are identified, the genetic variants could be used as tools in MAS analyses (Tong and Sun, 2015).

Considering common carp as a traditional cultivated species and commercial value, it has always been an important mission for breeders to cultivate new varieties with better characters. Up to now, QTL studies in common carp have covered a wide range of traits including growth, cold tolerance, meat quality, muscle fiber-related, sex determination, etc (Sun and Liang, 2004; Zhang et al., 2011; Laghari et al., 2013; Peng et al., 2016; Zheng et al., 2016). Compared to other economic traits, QTL analyses for FCE traits in aquaculture fish are rarely reported due to the difficulty to measure the phenotypes for these traits (Lu et al., 2017; Pang et al., 2017; Pang et al., 2018). So, it is expected that more genetics studies on FCE traits will be considered so that molecular breeding strategies for these traits would be developed. In this study, we constructed a high-density and high-resolution genetic linkage map for Songpu mirror carp and performed the first fine-scale QTL mapping for FCE. A group of closely linked markers and potential candidate genes were identified, which open new opportunities for MAS implementation, which can ultimately accelerate breeding mirror carp for high feed conversion efficiency.

Ultra-High Density Genetic Map for Mirror Carp

It is well known that the separation type of mapping population directly affects the efficiency of linkage map construction (Wei et al., 2019). The commonly used mapping families for the construction of linkage genetic maps include F_2 and backcross (BC) families, double haploid (DH), and recombinant inbred lines (RIL). However, it is a big challenge to construct DH or RIL families in most teleost fish, and the constructions of F_2 and BC families usually take a relatively long time (Peng et al., 2016). The F_1 progeny displays many different types of segregation. So, Grattapaglia and Sederoff proposed using the F_1 family as the mapping panel with a double pseudo-testcross strategy which has been successfully applied to genetic linkage map construction in many aquaculture species (Grattapaglia and Sederoff, 1994; Wu et al., 2010; Song et al., 2012; Peng et al., 2016). In this study, the parents of the F_1 population were derived from the cultured families of mirror carp at the Hulan station of the Heilongjiang Fisheries Research Institute of the Chinese Academy of Fishery Sciences, and the genetic distances among these fish were estimated using a panel of polymorphic microsatellite markers. A male and a female mature fish (F_0) which showed a relatively high genetic distance were used to generate an experimental family (F_1) by artificial crossing. Therefore, the mapping family used in this study with a double pseudo-testcross strategy was suitable for the construction of a genetic linkage map.

Obviously, a high-quality genetic map is an essential tool for QTL mapping with high efficiency and accuracy. However, a limited number of traditional markers make it difficult to cover the whole genome (Wen et al., 2020). The SNP marker is the most abundant and polymorphic marker in the genome which is very suitable for high-density linkage map construction. As an effective way, the second-generation sequencing technology makes it possible to obtain a sufficient number of SNP markers for linkage map construction in a short time. However, the call rate and genotyping accuracy of SNPs are critically important for the construction of a high-quality linkage map, as it is well known that missing data and genotyping errors would lead to incorrect map orders (Hackett and Broadfoot, 2003).

Compared with other SNP genotyping approaches, Affymetrix Axiom SNP genotyping platform has become an effective solution for large-scale SNP genotyping as it has higher call rates (>99%) and higher accuracy and has been successfully applied to the construction of the linkage map and QTL analysis of multiple species (Andreas et al., 2013; Sushma et al., 2016; Higgins et al., 2018; Gong et al., 2019; You et al., 2019). Recently, two high-density genetic linkage maps have been constructed based on Affymetrix Axiom SNP genotyping data for channel catfish (*Ictalurus punctatus*) and Yellow River carp (*C. carpio*), presenting the highest density linkage maps in aquaculture species (Li et al., 2015; Peng, et al., 2016). In this study, we also chose the same high-throughput SNP genotyping platform for SNP genotyping and constructed a high-density and high-accurate linkage genetic map for mirror carp. In order to get high-quality makers, we used more stringent data filtering (missing value < 1%) than the former studies of high-density linkage map construction. As a result, a total of 28,416 high-quality SNP markers were successfully mapped. The genetic map covered 99.07% of the genome with a density of 0.33 cM average distance, which demonstrates its power in the detection of potential QTL associated with FCE in mirror carp at a fine scale.

QTL Analysis and Candidate Genes

Most economically important traits of fish such as growth, disease resistance, and flesh quality are controlled by multiple genes, environmental factors, and their interactions (Song et al., 2012; Shao et al., 2015; Peng et al., 2016; Zheng et al., 2016). Phenotypic variation of these traits is thought to be caused by quantitative genetic variation that results from the segregation of alleles at multiple quantitative trait loci (QTL) (Mackay et al., 2009; Zhang et al., 2021). The purpose of QTL mapping is to understand the number and effect of genes that determine traits and to assist in the selection of breeding to accelerate the genetic improvement of important traits (Doerge, 2002; Naish and Hard 2008).

A number of studies looking for genetic mechanisms affecting feed efficiency, which involved numerous biological processes and functional pathways in livestock and poultry, have been reported using different methods (Sherman et al., 2009; Do et al., 2014; Ferreira et al., 2016; Silva et al., 2019; Li et al., 2020; Li et al., 2021). Compared to the livestock and poultry, it has been a challenge to map loci associated with feed efficiency because the phenotypic value of each individual is generally difficult to obtain

in aquaculture species (Lu et al., 2017; Pang et al., 2017). To date, only a few QTL analyses associated with feed conversion efficiency have been reported in aquaculture species. Preliminary QTL results about FCE in fish are from Liu's study, which used AFLP markers to construct a catfish genetic map and found a QTL associated with FCE (Liu, 2001). Zimmerman et al. revealed three QTL for the number of pyloric caeca in three LGs of rainbow trout, and this is an important index associated with FCE (Zimmerman et al., 2005; Tong and Sun, 2015). Recently, Lu et al. performed the QTL mapping of feed conversion rate (FCR) in mirror carp based on two mapping panels consisting of 92 and 68 samples and using 507 and 307 markers, respectively. As a result, 18 QTL affecting FCR were detected in two datasets (Lu et al., 2017). Pang et al. constructed a high-resolution genetic linkage map in a full-sib F_1 family of crucian carp (*Carassius auratus*) consisting of 113 progenies with 8,460 SNP markers. Eight FCE-related QTL and seven candidate genes involved in energy metabolism, digestion, biosynthesis, etc were detected (Pang et al., 2017). Although QTL about feed efficiency traits in common carp have been reported, not all of the genetic variants of the traits have been captured as the sample size of the mapping populations and the number of markers used were relatively small. QTL mapping has a large positioning range and there are too many genes in the location range. So the candidate genes were difficult to be determined. Historically, in the absence of high throughput genotyping data, it was difficult to perform QTL fine mappings and candidate gene identification. In this study, we were able to take advantage of the common carp 250K SNP genotyping array and construct an ultra-high density linkage map for mirror carp, providing new insights into the FCE trait and related genes. In the present investigation, a total of 17 QTL associated with FCE were mapped on four linkage groups, explaining 8.9–15.9% of the PVE, and nine candidate genes related to FCE involved in multiple biological processes were identified in common carp.

Feed conversion efficiency is a complex trait, which involves many physiological processes such as feeding, digestion, biosynthesis, metabolism, and so on, and is driven by a series of biological pathways. In our study, nine candidate genes associated with FCE were identified in common carp. Among these genes, *ACACA* has been reported as a part of a single multifunctional polypeptide in eukaryotes, and plays a critical role in the metabolism of fatty acid biosynthesis (Abu-Elheiga et al., 1995). *ACACA* is a biotin-containing enzyme that catalyzes the carboxylation of acetyl-CoA to malonyl-CoA, the rate-limiting step in fatty acid synthesis (Brusselmans et al., 2005). *ACACA* may induce dysregulation of lipid metabolism in human and mouse is known to result in metabolic diseases, such as obesity and diabetes (Wakil and Abu-Elheiga, 2009), it may also result in severe metabolic disorders in lactating cows (Loor et al., 2007). Therefore, we speculated that *ACACA* may be directly associated with fat metabolism and may subsequently affect FCE in common carp. The protein encod SCAF4 containing a domain, which found in hepatocyte growth factor-regulated tyrosine kinase substrate (Hrs), plays a critical role in the recycling and involved in endocytic trafficking (Li et al., 2002; Lea et al., 2019). So, we consider that its molecular function may be related to the growth of common carp, and may have a specific function related to FCE. *SLC2A5* has been reported as a facilitated glucose/fructose transmembrane transport which is the key solute carrier in the carbohydrate metabolic process of glucose and gluconeogenesis (Berghe, 1996). It takes part in carbohydrate digestion and absorption and leads to a significantly enhanced rate of triglyceride synthesis. *SLC2A5* is also responsible for fructose uptake by the small intestine in mammals (Wu et al., 2009; Nomura et al., 2016). High expression of *SLC2A5* in the small intestine in rats or mice leads to increased fructose absorption (Wright et al., 2007). A recent study showed that a higher expression level of *SLC2A5* was related to higher feed conversion efficiency-related traits (Diniz et al., 2020). Hence, we suggested that this gene was associated with food digestion, carbohydrate metabolic processing, and further influences on FCE in common carp. *TNMD* expression is highly affected by obesity, adipose tissue location, and weight loss in human adipose tissue, indicating that *TNMD* may play a role in adipose tissue function (Saiki et al., 2009). Therefore, we suggested that it may be associated with FCE by intervening in influence energy deposition and fat accumulation in common carp. The protein encoded by this gene has a domain associated with the catalytic domain of sugar utilizing enzymes, including maltooligosyl trehalose synthase, glycogen branching enzyme, glycogen debranching enzyme, isoamylase, etc, which endows PCDH1 with the capability to play several functions in different pathways (Aron et al., 2017). So, we suggested that it takes part in digestion and absorption and might be related to feed efficiency in common carp. FOXO is also involved in lipid metabolism and has been reported as one of the important candidate genes associated with obesity and body mass in human (Kim et al., 2006). The mRNA levels of FOXO isoforms in rat livers were altered in response to fasting and re-feeding, which suggests that the genes respond differently to nutritional and hormonal factors (Imae et al., 2003). Hence, we suggested that FOXO was associated with fat accumulation and growth in common carp. AGO1 encodes a member of the argonaute family of proteins which can affect cell proliferation, motility, and apoptosis in human (Parisi et al., 2011). Tang et al. (2020) report that AGO1 participates in a mechanism that controls adipose tissues, insulin sensitivity, and whole-body metabolic state. They found that when challenged with an obesity-inducing high-fat high-sucrose (HFHS) diet, AGO1-knockout mice displayed significantly lower body weight gain and lower body fat, improved insulin sensitivity, and enhanced energy expenditure. In human donors with obesity or type 2 diabetes mellitus, AGO1 is expressed at higher levels than in healthy controls, which also supports the role of this pathway (Tang et al., 2020). So, we indicate it is also possibly associated with digestive and metabolic functions in common carp. *FFAR3* is a G protein-coupled receptor that is activated by a major product of dietary fiber digestion, the short-chain fatty acids (*SCFAs*), which are an essential energy source and signaling molecules that regulate various cellular processes and physiological functions (Stoddart et al., 2008; Zaibi et al., 2010). Recent studies have shown that these receptors are involved in the lipid metabolism in

various tissues and play an important role in the absorption of nutrients in animal intestine (Hara et al., 2013; Ichimura et al., 2014; Lu et al., 2015; Meza-Cuenca et al., 2018). Therefore, *FFAR3* may be a promising candidate gene for FCE in common carp, and further functional approaches are necessary to validate. *ARID1A* has been reported to affect cholesterol synthesis as well as glycogen metabolism-related proteins levels in human and could play a significant role to help cell proliferation and inhibit cell apoptosis (Flores-Alcantar et al., 2011; Bosse et al., 2013; Goldman et al., 2016). *ARID1A* deletion in isolated hepatocytes directly leads to free fatty acid-induced lipid accumulation and insulin resistance in mice. These findings reveal a new mechanism underlying the role of *ARID1A* in glucose and lipid metabolism (Qu et al., 2019). *ARID1A* has also been reported to have a positive correlation of expression with the body mass trait in human (Keildson et al., 2014; Giri et al., 2018). Thus, we speculate that *ARID1A* plays similar roles in regulating lipid accumulation and body mass in common carp.

Obviously, the QTL and genes detected in this study need to be further verified for their functional relatedness to FCE in the future. These novel findings would be used for future genetic and genomic researches of FCE traits, thereafter providing essential markers of MAS breeding for the potential improvement of feed efficiency in mirror carp.

CONCLUSION

In conclusion, taking advantage of the common carp 250K SNP genotyping array, high throughput genotyping data were accurately and efficiently collected from a mirror carp mapping family and used to construct a high-density, high-resolution genetic linkage map for mirror carp. This map processes the highest marker density among all the constructed genetic maps in mirror carp. Based on this valuable genetic map, fine-scale QTL mapping of FCE was performed and candidate functional genes were also identified. These candidate genes were identified as functionally related to lipid metabolism, carbohydrate metabolism, energy deposition, fat accumulation, digestion, regulating growth, and cell proliferation and differentiation. The present high-density genetic map and mapping results provide a basis for further genetic research of feed conversion efficiency and facilitate future MAS breeding for the feed conversion efficiency trait in common carp.1

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/, GCF_018340385.

ETHICS STATEMENT

The animal study was reviewed and approved by the Committee for Animal Experiments of the Heilongjiang River Fisheries Research Institute, the Chinese Academy of Sciences, China.

AUTHOR CONTRIBUTIONS

The authors' responsibilities were as follows: XZ conceived the study and drafted the manuscript. PL performed the experiments and phenotype investigation. DC prepared the fish population. GH performed the manuscript revision. All authors read and approved the final manuscript.

FUNDING

This work was supported by the China Ministry of Science and Technology 863 Hi-Tech Research and Development Program (No. 2011AA100402), The Ministry of Agriculture and Rural Affairs of China, The Central Public-interest Scientific Institution Basal Research Fund (No. 2019XT0605), and The National Natural Science Foundation of China (No. 31872557).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.778487/full#supplementary-material

REFERENCES

Abu-Elheiga, L., Jayakumar, A., Baldini, A., Chirala, S. S., and Wakil, S. J. (1995). Human Acetyl-CoA Carboxylase: Characterization, Molecular Cloning, and Evidence for Two Isoforms. *Proc. Natl. Acad. Sci.* 92 (9), 4011–4015. doi:10.1073/pnas.92.9.4011

Andreas, K., Almas, A., Clarissa, B., Frances, T., Yu, L., Smith, S., et al. (2013). Development of a High Density 600KSNPgenotypingarray for Chicken. *BMC Genomics* 14 (1), 59. doi:10.1186/1471-2164-14-59

Aron, M., Yu, B., H, L., He, J., Christopher, L., Shennan, L., et al. (2017). CDD/SPARCLE: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Res.* 45 (D1), D200–D203. doi:10.1093/nar/gkw1129

Ashton, D. T., Ritchie, P. A., and Wellenreuther, M. (2017). Fifteen Years of Quantitative Trait Loci Studies in Fish: Challenges and Future Directions. *Mol. Ecol.* 26 (6), 1465–1476. doi:10.1111/mec.13965

Barendse, W., Reverter, A., Bunch, R. J., Harrison, B. E., Barris, W., and Thomas, M. B. (2007). A Validated Whole-Genome Association Study of Efficient Food Conversion in Cattle. *Genet* 176 (3), 1893–1905. doi:10.1534/genetics.107.072637

Barone, S., Fussell, S. L., Singh, A. K., Lucas, F., Xu, J., Kim, C., et al. (2009). Slc2a5 (Glut5) Is Essential for the Absorption of Fructose in the Intestine and Generation of Fructose-Induced Hypertension. *J. Biol. Chem.* 284 (8), 5056–5066. doi:10.1074/jbc.M808128200

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

Begli, H. E., Torshizi, R. V., Masoudi, A. A., Ehsani, A., and Jensen, J. (2016). Longitudinal Analysis of Body Weight, Feed Intake and Residual Feed Intake in F2 Chickens. *Livestock Sci.* 184, 28–34. doi:10.1016/j.livsci.2015.11.018

Bosse, T., ter Haar, N. T., Seeber, L. M., Diest, P. J. v., Hes, F. J., Vasen, H. F., et al. (2013). Loss of ARID1A Expression and its Relationship with PI3K-Akt Pathway Alterations, TP53 and Microsatellite Instability in Endometrial Cancer. *Mod. Pathol.* 26 (11), 1525–1535. doi:10.1038/modpathol.2013.96

Bostock, J., Mcandrew, B., Richards, R., Jauncey, K., Telfer, T., Lorenzen, K., et al. (2010). Aquaculture: Global Status and Trends. *Phil. Trans. R. Soc. B* 365 (1554), 2897–2912. doi:10.1098/rstb.2010.0170

Brusselmans, K., De Schrijver, E., Verhoeven, G., and Swinnen, J. V. (2005). RNA Interference-Mediated Silencing of theAcetyl-CoA-Carboxylase-αGene Induces Growth Inhibition and Apoptosis of Prostate Cancer Cells. *Cancer Res.* 65 (15), 6719–6725. doi:10.1158/0008-5472.CAN-05-0571

Cai, W., Casey, D. S., and Dekkers, J. C. M. (2008). Selection Response and Genetic Parameters for Residual Feed Intake in Yorkshire Swine1. *J. Anim. Sci.* 86 (2), 287–298. doi:10.2527/jas.2007-0396

Chakravarti, A., Lasher, L. K., and Reefer, J. E. (1991). A Maximum Likelihood Method for Estimating Genome Length Using Genetic Linkage Data. *Genet* 128 (1), 175–182. doi:10.1093/genetics/128.1.175

Christoffels, A., Bartfai, R., Srinivasan, H., Komen, H., and Orban, L. (2006). Comparative Genomics in Cyprinids: Common Carp ESTs Help the Annotation of the Zebrafish Genome. *BMC Bioinformatics* 7 (Suppl. 5), S2. doi:10.1186/1471-2105-7-S5-S2

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a Universal Tool for Annotation, Visualization and Analysis in Functional Genomics Research. *Bioinformatics* 21 (18), 3674–3676. doi:10.1093/bioinformatics/bti610

de Koning, D. J., Windsor, D., Hocking, P. M., Burt, D. W., Law, A., Haley, C. S., et al. (2003). Quantitative Trait Locus Detection in Commercial Broiler Lines Using Candidate Regions1. *J. Anim. Sci.* 81 (5), 1158–1165. doi:10.2527/2003.8151158x

Delpuech, E., Aliakbari, A., Labrune, Y., Fève, K., Billon, Y., Gilbert, H., et al. (2021). Identification of Genomic Regions Affecting Production Traits in Pigs Divergently Selected for Feed Efficiency. *Genet. Sel. Evol.* 53 (1), 49. doi:10.1186/s12711-021-00642-1

Diniz, W. J. S., da Rosa, K. O., Tizioto, P. C., Mourão, G. B., de Oliveirae, P. S. N., de Souza, M. M., et al. (2020). FABP1 and SLC2A5 Expression Levels Affect Feed Efficiency-Related Traits. *Agri. Gene* 15, 100100. doi:10.1016/j.aggene.2019.100100

Do, D. N., Strathe, A. B., Ostersen, T., Pant, S. D., and Kadarmideen, H. N. (2014). Genome-wide Association and Pathway Analysis of Feed Efficiency in Pigs Reveal Candidate Genes and Pathways for Residual Feed Intake. *Front. Genet.* 5, 307. doi:10.3389/fgene.2014.00307

Doerge, R. W. (2002). Mapping and Analysis of Quantitative Trait Loci in Experimental Populations. *Nat. Rev. Genet.* 3, 43–52. doi:10.1038/nrg703

Easterling, W. E. (2007). Climate Change and the Adequacy of Food and Timber in the 21st century. *Proc. Natl. Acad. Sci.* 104 (50), 19679. doi:10.1073/pnas.0710388104

Ferreira, O., Zerlotti, M., Gonalves, C., Helena, B., Martins, B., Galvo, D., et al. (2016). Genomic Regions Associated with Feed Efficiency Indicator Traits in an Experimental Nellore Cattle Population. *Plos One* 11 (10), e0164390. doi:10.1371/journal.pone.0164390

Flores-Alcantar, A., Gonzalez-Sandoval, A., Escalante-Alcalde, D., and Lomelí, H. (2011). Dynamics of Expression of ARID1A and ARID1B Subunits in Mouse Embryos and in Cells during the Cell Cycle. *Cell Tissue Res* 345 (1), 137–148. doi:10.1007/s00441-011-1182-x

Fu, L., Jiang, Y., Wang, C., Mei, M., Zhou, Z., Jiang, Y., et al. (2020). A Genome-wide Association Study on Feed Efficiency Related Traits in Landrace Pigs. *Front. Genet.* 11, 692. doi:10.3389/fgene.2020.00692

Giri, A. K., Parekatt, V., Dwivedi, O. P., Banerjee, P., Bandesh, K., Prasad, G., et al. (2018). Common Variants of ARID1A and KAT2B Are Associated with Obesity in Indian Adolescents. *Sci. Rep.* 8 (1), 3964. doi:10.1038/s41598-018-22231-x

Gjedrem, T., and Baranski, M. (2009). *Selective Breeding in Aquaculture: An Introduction*. New York: Springer.

Goldman, A. R., Bitler, B. G., Schug, Z., Conejo-Garcia, J. R., Zhang, R., Speicher, D. W., et al. (2016). The Primary Effect on the Proteome of ARID1A-Mutated Ovarian Clear Cell Carcinoma Is Downregulation of the Mevalonate Pathway at the Post-Transcriptional Level. *Mol. Cell Proteomics* 15 (11), 3348–3360. doi:10.1074/mcp.m116.062539

Gong, H., Xiao, S., Li, W., Huang, T., Huang, X., Yan, G., et al. (2019). Unravelling the Genetic Loci for Growth and Carcass Traits in Chinese Bamaxiang Pigs Based on a 1.4 Million SNP Array. *J. Anim. Breed. Genet.* 136 (1), 3–14. doi:10.1111/jbg.12365

Grattapaglia, D., and Sederoff, R. (1994). Genetic Linkage Maps of Eucalyptus Grandis and Eucalyptus Urophylla Using a Pseudo-testcross: Mapping Strategy and RAPD Markers. *Genet* 137 (4), 1121–1137. doi:10.1093/genetics/137.4.1121

Hackett, C. A., and Broadfoot, L. B. (2003). Effects of Genotyping Errors, Missing Values and Segregation Distortion in Molecular Marker Data on the Construction of Linkage Maps. *Heredity* 90 (1), 33–38. doi:10.1038/sj.hdy.6800173

Hara, T., Kimura, I., Inoue, D., Ichimura, A., and Hirasawa, A. (2013). Free Fatty Acid Receptors and Their Role in Regulation of Energy Metabolism. *Rev. Physiol. Biochem. Pharmacol.* 164, 77–116. doi:10.1007/112_2013_13

Herd, R., Archer, J., and Arthur, P. (2003). Reducing the Cost of Beef Production through Genetic Improvement in Residual Feed Intake: Opportunity and Challenges to Application. *J. Anim. Sci.* 81 (13), 9–17. doi:10.2527/2003.8113_suppl_1E9x

Higgins, M. G., Fitzsimons, C., McClure, M. C., McKenna, C., Conroy, S., Kenny, D. A., et al. (2018). GWAS and eQTL Analysis Identifies a SNP Associated with Both Residual Feed Intake and GFRA2 Expression in Beef Cattle. *Sci. Rep.* 8 (1), 14301. doi:10.1038/s41598-018-32374-6

Huo, X., Liu, S., Shao, T., Hua, H., Kong, Q., Wang, J., et al. (2014). GSK3 Protein Positively Regulates Type I Insulin-like Growth Factor Receptor through Forkhead Transcription Factors FOXO1/3/4. *J. Biol. Chem.* 289 (36), 24759–24770. doi:10.1074/jbc.M114.580738

Ichimura, A., Hasegawa, S., Kasubuchi, M., and Kimura, I. (2014). Free Fatty Acid Receptors as Therapeutic Targets for the Treatment of Diabetes. *Front. Pharmacol.* 5, 236–6. doi:10.3389/fphar.2014.00236

Imae, M., Fu, Z., Yoshida, A., Noguchi, T., and Kato, H. (2003). Nutritional and Hormonal Factors Control the Gene Expression of FoxOs, the Mammalian Homologues of DAF-16. *J. Mol. Endocrinol.* 30 (2), 253–262. doi:10.1677/jme.0.0300253

Jia, Z., Chen, L., Ge, Y., Li, S., Peng, W., Li, C., et al. (2020). Genetic Mapping of Koi Herpesvirus Resistance (KHVR) in Mirror Carp (*Cyprinus carpio*) Revealed Genes and Molecular Mechanisms of Disease Resistance. *Aquaculture* 519, 734850. doi:10.1016/j.aquaculture.2019.734850

Jun Wang, J. (2012). Genetic Map Construction and Quantitative Trait Locus (QTL) Analysis on Growth-Related Traits in Common Carp (*Cyprinus carpio* L.). *Afr. J. Biotechnol.* 11 (31), 7875–7884. doi:10.5897/ajb11.3268

Keildson, S., Fadista, J., Ladenvall, C., Hedman, A. K., Elgzyri, T., Small, K. S., et al. (2014). Expression of Phosphofructokinase in Skeletal Muscle Is Influenced by Genetic Variation and Associated with Insulin Sensitivity. *Diabetes* 63 (3), 1154–1165. doi:10.2337/db13-1301

Kim, J.-R., Jung, H. S., Bae, S.-W., Kim, J. H., Park, B. L., Choi, Y. H., et al. (2006). Polymorphisms in FOXO Gene Family and Association Analysis with BMI*. *Obesity* 14 (2), 188–193. doi:10.1038/oby.2006.24

Kuang, Y., Zheng, X., Lv, W., Cao, D., and Sun, X. (2015). Mapping Quantitative Trait Loci for Flesh Fat Content in Common Carp (*Cyprinus carpio*). *Aquaculture* 435, 100–105. doi:10.1016/j.aquaculture.2014.09.020

Laghari, M. Y., Zhang, Y., Lashari, P., Zhang, X., Xu, P., Xin, B., et al. (2013). Quantitative Trait Loci (QTL) Associated with Growth Rate Trait in Common Carp (*Cyprinus carpio*). *Aquacult Int.* 21 (6), 1373–1379. doi:10.1007/s10499-013-9639-4

Lande, R., and Thompson, R. (1990). Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. *Genet* 124 (3), 743–756. doi:10.1093/genetics/124.3.743

Lea, H., Richard, M., Alejandro, P., Takayuki, N., Nicholas, J., Reuven, A., et al. (2019). SCAF4andSCAF8,mRNAAnti-TerminatorProteins. *Cell* 177 (7), 1797–1813. doi:10.1016/j.cell.2019.04.038

Li, G.-S., Liu, W.-W., Zhang, F., Zhu, F., Yang, F.-X., Hao, J.-P., et al. (2020). Genome-wide Association Study of Bone Quality and Feed Efficiency-Related Traits in pekin Ducks. *Genomics* 112 (6), 5021–5028. doi:10.1016/j.ygeno.2020.09.023

Li, H., Ye, G., and Wang, J. (2007). A Modified Algorithm for the Improvement of Composite Interval Mapping. *Genet* 175 (1), 361–374. doi:10.1534/genetics.106.066811

Li, W., Zheng, M., Zhao, G., Wang, J., Liu, J., Wang, S., et al. (2021). Identification of QTL Regions and Candidate Genes for Growth and Feed Efficiency in Broilers. *Genet. Sel. Evol.* 53 (1), 13. doi:10.1186/s12711-021-00608-3

Li, Y., Chin, L.-S., Levey, A. I., and Li, L. (2002). Huntingtin-associated Protein 1 Interacts with Hepatocyte Growth Factor-Regulated Tyrosine Kinase Substrate and Functions in Endosomal Trafficking. *J. Biol. Chem.* 277 (31), 28212–28221. doi:10.1074/jbc.m111612200

Li, Y., Liu, S., Qin, Z., Waldbieser, G., Wang, R., Sun, L., et al. (2015). Construction of a High-Density, High-Resolution Genetic Map and its Integration with BAC-Based Physical Map in Channel Catfish. *DNA Res.* 22 (1), 39–52. doi:10.1093/dnares/dsu038

Liu, Z. (2001). "Gene Mapping, Marker-Assisted Selection, Gene Cloning, Genetic Engineering and Integrated Genetic Improvement Programs at Auburn University,". *Fish Genetics Research in Member Countries and Institutions of the International Network on Genetics in Aquaculture*. Editors M. V. Gupta and B. O. Acosta, 64, 179.

Loor, J. J., Everts, R. E., Bionaz, M., Dann, H. M., Morin, D. E., Oliveira, R., et al. (2007). Nutrition-induced Ketosis Alters Metabolic and Signaling Gene Networks in Liver of Periparturient Dairy Cows. *Physiol. Genomics* 32 (1), 105–116. doi:10.1152/physiolgenomics.00188.2007

Lu, C., Laghari, M. Y., Zheng, X., Cao, D., Zhang, X., Kuang, Y., et al. (2017). Mapping Quantitative Trait Loci and Identifying Candidate Genes Affecting Feed Conversion Ratio Based onto Two Linkage Maps in Common Carp (*Cyprinus carpio* L). *Aquaculture* 468, 585–596. doi:10.1016/j.aquaculture.2016.10.040

Lu, Z., Gui, H., Yao, L., Yan, L., Martens, H., Aschenbach, J. R., et al. (2015). Short-chain Fatty Acids and Acidic pH Upregulate UT-B, GPR41, and GPR4 in Rumen Epithelial Cells of Goats. *Am. J. Physiology-Regulatory, Integr. Comp. Physiol.* 308 (4), R283–R293. doi:10.1152/ajpregu.00323.2014

Lv, W., Zheng, X., Kuang, Y., Cao, D., Yan, Y., and Sun, X. (2016). QTL Variations for Growth-Related Traits in Eight Distinct Families of Common Carp (*Cyprinus carpio*). *BMC Genet.* 17 (1), 65. doi:10.1186/s12863-016-0370-9

Mackay, T. F. C., Stone, E. A., and Ayroles, J. F. (2009). The Genetics of Quantitative Traits: Challenges and Prospects. *Nat. Rev. Genet.* 10 (8), 565–577. doi:10.1038/nrg2612

Meza-Cuenca, F., Medina-Contreras, J. M. L., Mailloux-Salinas, P., Bautista-Hernández, L. A., Buentello-Volante, B., Domínguez-López, A., et al. (2018). Characterization of Free Fatty Acid Receptors Expression in an Obesity Rat Model with High Sucrose Diet. *J. Receptors Signal Transduction* 38 (1), 76–82. doi:10.1080/10799893.2018.1426609

Miao, Y., Mei, Q., Fu, C., Liao, M., Liu, Y., Xu, X., et al. (2021). Genome-wide Association and Transcriptome Studies Identify Candidate Genes and Pathways for Feed Conversion Ratio in Pigs. *BMC genomics* 22 (1), 294. doi:10.1186/s12864-021-07570-w

Mignon-Grasteau, S., Rideau, N., Gabriel, I., Chantry-Darmon, C., Boscher, M. Y., Sellier, N., et al. (2015). Detection of QTL Controlling Feed Efficiency and Excretion in Chickens Fed a Wheat-Based Diet. *Genet. Sel. Evol.* 47 (1), 74–13. doi:10.1186/s12711-015-0156-y

Naish, K. A., and Hard, J. J. (2008). Bridging the gap between the Genotype and the Phenotype: Linking Genetic Variation, Selection and Adaptation in Fishes. *Fish and Fisheries* 9, 396–422. doi:10.1111/j.1467-2979.2008.00302.x

Nomura, N., Verdon, G., Kang, H. J., Shimamura, T., Nomura, Y., Sonoda, Y., et al. (2015). Structure and Mechanism of the Mammalian Fructose Transporter GLUT5. *Nature* 526 (7573), 397–401. doi:10.1038/nature14909

Pang, M., Fu, B., Yu, X., Liu, H., Wang, X., Yin, Z., et al. (2017). Quantitative Trait Loci Mapping for Feed Conversion Efficiency in Crucian Carp (*Carassius auratus*). *Sci. Rep.* 7 (1), 16971. doi:10.1038/s41598-017-17269-2

Pang, M., Luo, W., Fu, B., Yu, X., Zhou, Y., and Tong, J. (2018). Transcriptomic Profiles of Brain Provide Insights into Molecular Mechanism of Feed Conversion Efficiency in Crucian Carp (*Cyprinus carpio*). *Int. J. Mol. Sci.* 19 (3), 858. doi:10.3390/ijms19030858

Parisi, C., Giorgi, C., Batassa, E. M., Braccini, L., Maresca, G., D'agnano, I., et al. (2011). Ago1 and Ago2 Differentially Affect Cell Proliferation, Motility and Apoptosis when Overexpressed in SH-Sy5y Neuroblastoma Cells. *FEBS Lett.* 585 (19), 2965–2971. doi:10.1016/j.febslet.2011.08.003

Peng, W., Xu, J., Zhang, Y., Feng, J., Dong, C., Jiang, L., et al. (2016). An Ultra-high Density Linkage Map and QTL Mapping for Sex and Growth-Related Traits of Common Carp (*Cyprinus carpio*). *Sci. Rep.* 6, 26693. doi:10.1038/srep26693

Qu, Y.-L., Deng, C.-H., Luo, Q., Shang, X.-Y., Wu, J.-X., Shi, Y., et al. (2019). Arid1a Regulates Insulin Sensitivity and Lipid Metabolism. *EBioMedicine* 42, 481–493. doi:10.1016/j.ebiom.2019.03.021

Rezende, F. M., Ferraz, J. B. S., Eler, J. P., Silva, R. C. G., Mattos, E. C., and Ibáñez-Escriche, N. (2012). Study of Using Marker Assisted Selection on a Beef Cattle Breeding Program by Model Comparison. *Livestock Sci.* 147, 40–48. doi:10.1016/j.livsci.2012.03.017

Rice, J. C., and Garcia, S. M. (2011). Fisheries, Food Security, Climate Change, and Biodiversity: Characteristics of the Sector and Perspectives on Emerging Issues. *Ices J. Mar. Sci.* 68 (6), 1343–1353. doi:10.1093/icesjms/fsr041

Saiki, A., Olsson, M., Jerna˚s, M., Gummesson, A., McTernan, P. G., Andersson, J., et al. (2009). Tenomodulin Is Highly Expressed in Adipose Tissue, Increased in Obesity, and Down-Regulated during Diet-Induced Weight Loss. *J. Clin. Endocrinol. Metab.* 94 (10), 3987–3994. doi:10.1210/jc.2009-0292

Shao, C., Niu, Y., Rastas, P., Liu, Y., Xie, Z., Li, H., et al. (2015). Genome-wide SNP Identification for the Construction of a High-Resolution Genetic Map of Japanese Flounder (P*aralichthys olivaceus)*: Applications to QTL Mapping of Vibrio Anguillarum Disease Resistance and Comparative Genomic Analysis. *DNA Res.* 22 (2), 161–170. doi:10.1093/dnares/dsv001

Sherman, E. L., Nkrumah, J. D., Li, C., Bartusiak, R., Murdoch, B., and Moore, S. S. (2009). Fine Mapping Quantitative Trait Loci for Feed Intake and Feed Efficiency in Beef Cattle1. *J. Anim. Sci.* 87 (1), 37–45. doi:10.2527/jas.2008-0876

Silva, É. F., Lopes, M. S., Lopes, P. S., and Gasparino, E. (2019). A Genome-wide Association Study for Feed Efficiency-Related Traits in a Crossbred Pig Population. *Animal* 13 (11), 2447–2456. doi:10.1017/S1751731119000910

Song, W., Li, Y., Zhao, Y., Liu, Y., Niu, Y., Pang, R., et al. (2012). Construction of a High-Density Microsatellite Genetic Linkage Map and Mapping of Sexual and Growth-Related Traits in Half-Smooth Tongue Sole (*Cynoglossus Semilaevis*). *Plos One* 7 (12), e52097. doi:10.1371/journal.pone.0052097

Stoddart, L. A., Smith, N. J., and Milligan, G. (2008). International Union of Pharmacology. LXXI. Free Fatty Acid Receptors FFA1, -2, and -3: Pharmacology and Pathophysiological Functions. *Pharmacol. Rev.* 60 (4), 405–417. doi:10.1124/pr.108.00802

Su, S., Raouf, B., He, X., Cai, N., Li, X., Yu, J., et al. (2020). Genome Wide Analysis for Growth at Two Growth Stages in A New Fast-Growing Common Carp Strain (*Cyprinus carpio* L.). *Sci. Rep.* 10 (1), 7259. doi:10.1038/s41598-020-64037-w

Sun, X., and Liang, L. (2004). A Genetic Linkage Map of Common Carp (*Cyprinus carpio* L.) and Mapping of a Locus Associated with Cold Tolerance. *Aquaculture* 238 (1-4), 165–172. doi:10.1016/S0044-8486(03)00445-9

Sushma, T., Krishnamurthy, S., Kumar, V., Singh, B., Rao, A., Mithra, S., et al. (2016). Mapping QTLsQTL for Salt Tolerance inRice(Oryzasativa L.)by Bulked Segregant Analysis of Recombinant Inbred Lines Using 50K SNPChip. *PLoS ONE* 11 (4), 1–19. doi:10.1371/journal.pone.0153610

Tang, X., Miao, Y., Luo, Y., Sriram, K., Qi, Z., Lin, F.-M., et al. (2020). Suppression of Endothelial AGO1 Promotes Adipose Tissue Browning and Improves Metabolic Dysfunction. *Circulation* 142 (4), 365–379. doi:10.1161/circulationaha.119.041231

Tolppanen, A.-M., Pulkkinen, L., Kolehmainen, M., Schwab, U., Lindström, J., Tuomilehto, J., et al. (2007). Tenomodulin Is Associated with Obesity and Diabetes Risk: The Finnish Diabetes Prevention Study*. *Obesity* 15 (5), 1082–1088. doi:10.1038/oby.2007.613

Tong, J., and Sun, X. (2015). Genetic and Genomic Analyses for Economically Important Traits and Their Applications in Molecular Breeding of Cultured Fish. *Sci. China Life Sci.* 58 (2), 178–186. doi:10.1007/s11427-015-4804-9

van den Berghe, G. (1996). Disorders of Gluconeogenesis. *J. Inherit. Metab. Dis.* 19 (4), 470–477. doi:10.1007/BF01799108

Van Ooijen, J. (2006). *JoinMap®4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen, Netherlands: Kyazma, BV.

Van Ooijen, J. (2004). *MAPQTL5.0, Software for the Mapping of Quantitative Trait Loci in Experimental Population*. Wageningen, Netherlands: Plant Research International.

Voorrips, R. E. (2002). MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *J. Hered.* 93 (1), 77–78. doi:10.1093/jhered/93.1.77

Wakil, S. J., and Abu-Elheiga, L. A. (2009). Fatty Acid Metabolism: Target for Metabolic Syndrome. *J. Lipid Res.* 50 (Suppl. l), S138–S143. doi:10.1194/jlr.r800079-jlr200

Wang, B., Zhu, J., Mounzih, K., Chehab, E. F., Ke, Y., and Chehab, F. F. (2009). Overexpression of the Transcription Factor Foxo4 Is Associated with Rapid Glucose Clearance. *Mol. Cell. Endocrinol.* 307 (1-2), 217–223. doi:10.1016/j.mce.2009.04.011

Wang, X., Zhang, X., Li, W., Zhang, T., Li, C., and Sun, X. (2012b). Mapping and Genetic Effect Analysis on Quantitative Trait Loci Related to Feed Conversion Ratio of Common Carp (In Chinese). *Acta Hydrobiol. Sin.* 36 (2), 177–196. doi:10.3724/SP.J.1035.2012.00177

Watamoto, Y., Futawaka, K., Hayashi, M., Matsushita, M., Mitsutani, M., Murakami, K., et al. (2019). Insulin-like Growth Factor-1 Directly Mediates Expression of Mitochondrial Uncoupling Protein 3 via Forkhead Box O4. *Growth Horm. IGF Res.* 46-47, 24–35. doi:10.1016/j.ghir.2019.05.003

Wei, J., Chen, Y., and Wang, W. (2019). A High-Density Genetic Linkage Map and QTL Mapping for Sex and Growth-Related Traits of Large-Scale Loach (*Paramisgurnus dabryanus*). *Front. Genet.* 10, 1023. doi:10.3389/fgene.2019.01023

Wen, Y., Fang, Y., Hu, P., Tan, Y., Wang, Y., Hou, L., et al. (2020). Construction of a High-Density Genetic Map Based on SLAF Markers and QTL Analysis of Leaf Size in Rice. *Front. Plant Sci.* 11, 1143. doi:10.3389/fgene.2020.0114310.3389/fpls.2020.01143

Williams, D. R., Li, W., Hughes, M. A., Gonzalez, S. F., Vernon, C., Vidal, M. C., et al. (2008). Genomic Resources and Microarrays for the Common Carp *Cyprinus carpio* L. *J. Fish. Biol.* 72 (9), 2095–2117. doi:10.1111/j.1095-8649.2008.01875.x

Wright, E. M., Hirayama, B. A., and Loo, D. F. (2007). Active Sugar Transport in Health and Disease. *J. Intern. Med.* 261 (1), 32–43. doi:10.1111/j.1365-2796.2006.01746.x

Wringe, B. F., Devlin, R. H., Ferguson, M. M., Moghadam, H. K., Sakhrani, D., and Danzmann, R. G. (2010). Growth-related Quantitative Trait Loci in Domestic and Wild Rainbow trout *(Oncorhynchus mykiss). BMC Genet.* 11, 63. doi:10.1186/1471-2156-11-63

Wu, S., Yang, J., Huang, Y., Li, Y., Yin, T., Wullschleger, S. D., et al. (20102010). An Improved Approach for Mapping Quantitative Trait Loci in a Pseudo-testcross: Revisiting a poplar Mapping Study. *Bioinform. Biol. Insights* 4 (4), 1–8. doi:10.4137/bbi.s4153

Xu, J., Zhao, Z., Zhang, X., Zheng, X., Li, J., Jiang, Y., et al. (2014a). Development and Evaluation of the First High-Throughput SNP Array for Common Carp (*Cyprinus carpio*). *BMC Genomics* 15 (1), 307. doi:10.1186/1471-2164-15-307

Xu, P., Xu, J., Liu, G., Chen, L., Zhou, Z., Peng, W., et al. (2019). The Allotetraploid Origin and Asymmetrical Genome Evolution of the Common Carp *Cyprinus carpio*. *Nat. Commun.* 10 (1), 4625. doi:10.1038/s41467-019-12644-1

Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., et al. (2014b). Genome Sequence and Genetic Diversity of the Common Carp, *Cyprinus carpio*. *Nat. Genet.* 46 (11), 1212–1219. doi:10.1038/ng.3098

Ye, S., Chen, Z.-T., Zheng, R., Diao, S., Teng, J., Yuan, X., et al. (2020). New Insights from Imputed Whole-Genome Sequence-Based Genome-wide Association Analysis and Transcriptome Analysis: The Genetic Mechanisms Underlying Residual Feed Intake in Chickens. *Front. Genet.* 11, 243. doi:10.3389/fgene.2020.00243

You, Q., Yang, X., Peng, Z., Islam, M. S., Sood, S., Luo, Z., et al. (2019). Development of an Axiom Sugarcane100K SNP Array for Genetic Map Construction and QTL Identification. *Theor. Appl. Genet.* 132 (10), 2829–2845. doi:10.1007/s00122-019-03391-4

Yue, G. H. (2014). Recent Advances of Genome Mapping and Marker-Assisted Selection in Aquaculture. *Fish Fish* 15 (3), 376–396. doi:10.1111/faf.12020

Zaibi, M. S., Stocker, C. J., O'Dowd, J., Davies, A., Bellahcene, M., Cawthorne, M. A., et al. (2010). Roles of GPR41 and GPR43 in Leptin Secretory Responses of Murine Adipocytes to Short Chain Fatty Acids. *FEBS Lett.* 584 (11), 2381–2386. doi:10.1016/j.febslet.2010.04.027

Zhang, X., Li, C., Lu, C., and Cao, D. (2021). Elite Allele Mining for Growth Rate Traits in Common Carp (*Cyprinus carpio*) by Association Analysis. *Aquac. Res.* 52 (3), 1192–1200. doi:10.1111/are.14977

Zhang, Y., Xu, P., Lu, C., Kuang, Y., Zhang, X., Cao, D., et al. (2011). Genetic Linkage Mapping and Analysis of Muscle Fiber-Related QTLs in Common Carp (*Cyprinus carpio* L.). *Mar. Biotechnol.* 13 (3), 376–392. doi:10.1007/s10126-010-9307-x

Zheng, X., Kuang, Y., Lv, W., Cao, D., Sun, Z., and Sun, X. (2016). Genome-wide Association Study for Muscle Fat Content and Abdominal Fat Traits in Common Carp (*Cyprinus carpio*). *Plos One* 11 (12), e0169127. doi:10.1371/journal.pone.0169127

Zhu, J., Mounzih, K., Chehab, E. F., Mitro, N., Saez, E., and Chehab, F. F. (2010). Effects of FoxO4 Overexpression on Cholesterol Biosynthesis, Triacylglycerol Accumulation, and Glucose Uptake. *J. Lipid Res.* 51 (6), 1312–1324. doi:10.1194/jlr.M001586

Zimmerman, A., Wheeler, P., Ristow, S., and Thorgaard, G. (2005). Composite Interval Mapping Reveals Three QTL Associated with Pyloric Caeca Number in Rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* 247 (1-4), 85–95. doi:10.1016/j.aquaculture.2005.02.029

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 in Genetics

ORIGINAL RESEARCH
published: 26 November 2021
doi: 10.3389/fgene.2021.772442

Comparison of Gene Expression Between Resistant and Susceptible Families Against VP_{AHPND} and Identification of Biomarkers Used for Resistance Evaluation in *Litopenaeus vannamei*

Qian Zhang[1,2], *Yang Yu*[1,3]*, *Zheng Luo*[1,2], *Jianhai Xiang*[1,3] and *Fuhua Li*[1,3,4,5]*

[1]Key Laboratory of Experimental Marine Biology, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, [4]Center for Ocean Mega-Science, Chinese Academy of Sciences, Qingdao, China, [5]The Innovation of Seed Design, Chinese Academy of Sciences, Wuhan, China

OPEN ACCESS

Edited by:
Li Zhou,
Institute of Hydrobiology (CAS), China

Reviewed by:
Zhiyi Bai,
Shanghai Ocean University, China
Mukunda Goswami,
Central Institute of Fisheries Education (ICAR), India

***Correspondence:**
Yang Yu
yuyang@qdio.ac.cn
Fuhua Li
fhli@qdio.ac.cn

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

Received: *08 September 2021*
Accepted: *01 November 2021*
Published: *26 November 2021*

Citation:
Zhang Q, Yu Y, Luo Z, Xiang J and Li F (2021) Comparison of Gene Expression Between Resistant and Susceptible Families Against VP_{AHPND} and Identification of Biomarkers Used for Resistance Evaluation in Litopenaeus vannamei. *Front. Genet. 12:772442. doi: 10.3389/fgene.2021.772442*

Acute hepatopancreatic necrosis disease (AHPND) has caused a heavy loss to shrimp aquaculture since its outbreak. *Vibrio parahaemolyticus* (VP_{AHPND}) is regarded as one of the main pathogens that caused AHPND in the Pacific white shrimp *Litopenaeus vannamei*. In order to learn more about the mechanism of resistance to AHPND, the resistant and susceptible shrimp families were obtained through genetic breeding, and comparative transcriptome approach was used to analyze the gene expression patterns between resistant and susceptible families. A total of 95 families were subjected to VP_{AHPND} challenge test, and significant variations in the resistance of these families were observed. Three pairs of resistant and susceptible families were selected for transcriptome sequencing. A total of 489 differentially expressed genes (DEGs) that presented in at least two pairwise comparisons were screened, including 196 DEGs highly expressed in the susceptible families and 293 DEGs in the resistant families. Among these DEGs, 16 genes demonstrated significant difference in all three pairwise comparisons. Gene set enrichment analysis (GSEA) of all 27,331 expressed genes indicated that some energy metabolism processes were enriched in the resistant families, while signal transduction and immune system were enriched in the susceptible families. A total of 32 DEGs were further confirmed in the offspring of the detected families, among which 19 genes were successfully verified. The identified genes in this study will be useful for clarifying the genetic mechanism of shrimp resistance against *Vibrio* and will further provide molecular markers for evaluating the disease resistance of shrimp in the breeding program.

Keywords: gene expression, disease resistance, molecular marker, *Vibrio parahaemolyticus*, *Litopenaeus vannamei*

INTRODUCTION

Litopenaeus vannamei is a commercially important aquaculture species, making up about 85% of total shrimp production in China (Qin et al., 2018). However, the shrimp aquaculture industry is continuously affected by the outbreak of viral and bacterial diseases, which have caused mass mortality and considerable economic losses. *Vibrio parahaemolyticus* carrying the PirA and PirB toxin genes in its plasmid (VP_{AHPND}) is one of the most destructive pathogens in shrimp aquaculture, causing acute hepatopancreatic necrosis disease (AHPND) or early mortality syndrome (EMS) in *L. vannamei* (Lee et al., 2015). VP_{AHPND} also caused AHPND in *Penaeus monodon* and *Exopalaemon carinicauda* (Soonthornchai et al., 2016; Ge et al., 2018). Therefore, prevention and control of AHPND are urgently needed in shrimp aquaculture.

Genetic selective breeding of disease resistance broodstock is a feasible and sustainable approach for the disease control. It has been proved to be efficient in controlling Taura syndrome virus (TSV). A disease-resistant line of *L. vannamei* against TSV had been established, and 18.4% increase in survival rate against TSV infection was obtained after one generation selection (Argue et al., 2002). For the selection of White spot syndrome virus (WSSV) resistance in *L. vannamei*, it showed that the average survival rates of generations G2 to G5 were 5.57%, 7.78%, 9.52%, and 13.79%, respectively (Huang et al., 2012). After three successive generation selection, the survival rates of *E. carinicauda* to VP_{AHPND} increased from 26.67% to 36.67% (Ge et al., 2018).

With the development of molecular biology, genomics approach offers a new possibility for accelerating the genetic selection process (Zhang X. et al., 2019). Identification of the major genes associated with disease resistance is the first step for marker-assisted selection (MAS) or gene-assisted selection (GAS) (Arora et al., 2019). So far, several genes associated with disease resistance have been reported in aquatic animals. Polymorphism of *LvALF* and *TRAF6* was reported to be associated with the resistance to WSSV in shrimp (Wang et al., 2011; Liu et al., 2014b; Liu et al., 2014a; Yu et al., 2017; Zhang Q. et al., 2019). A major quantitative trait locus (QTL) for the resistance of Atlantic salmon (*Salmo salar*) against infectious pancreatic necrosis (IPN) was discovered, which has been already applied in marker-assisted breeding of IPN-resistant fish (Houston et al., 2008; Moen et al., 2009; Woldemariam et al., 2020). However, knowledge about the resistance to AHPND in shrimp is still poorly understood. A comparison of individuals or families with significant phenotype difference by transcriptome sequencing is an efficient way for screening the trait-associated genes. Based on the transcriptome data, myosin, myosin heavy chain, and chitinase were proved to be related to growth performance in *L. vannamei* (Santos et al., 2018). Transcriptome comparison between the families with high growth rate and low growth rate also illustrated that the genes related to cuticle, chitin, and muscle proteins were upregulated exclusively in higher growth families (Santos et al., 2021). Besides, the gene profiles of the *Vibrio*-resistant and *Vibrio*-susceptible *Meretrix petechialis* families were analyzed, and several genes such as *Big-Def*, *CTL9*, and *Bax* were identified as candidate resistance-associated genes (Jiang et al., 2017).

In our previous work, we have carried out systematic family selection for the resistance trait to VP_{AHPND} in *L. vannamei*. A total of 95 families were subjected to VP_{AHPND} challenge test, and the results showed that significant resistance variations existed between different families. In the present study, we selected three AHPND resistant and three susceptible families for transcriptome sequencing and explored the differentially expressed genes (DEGs) in resistant and susceptible families. Several genes related to the resistance of shrimp against AHPND were identified. These genes might be developed as effective molecular markers for evaluating the disease resistance of shrimp, which could facilitate molecular marker-assisted breeding of shrimp.

MATERIALS AND METHODS

Selection Resistant and Susceptible Families Against Acute Hepatopancreatic Necrosis Disease

The full-sib families of *L. vannamei* were produced and stocked separately in Hainan Grand Suntop Ocean Breeding Co., Ltd, Wenchang, China. In order to identify the resistance of AHPND, the shrimp families were challenged by VP_{AHPND} each year, and the families with high survival rate were mated to generate the next-generation families. In 2018, 95 full-sib families with an average body weight of 2.10 g were selected for evaluation of resistance. For the challenge experiment, VP_{AHPND} was prepared according to the method described by Zhang Q. et al. (2019). About 100 healthy shrimp from each family were subjected to VP_{AHPND} immersion infection. Before the experiment, the shrimp were kept in the aquarium at a temperature of 26°C ± 1°C for 3 days to acclimate them to the culture conditions. Then, the VP_{AHPND} were added to the aquarium to make the concentration of VP_{AHPND} as 5×10^6 CFU/ml. Then, the dead shrimp were checked every 2 h. The mortality of each family was recorded for 72 h. The survival rates of the tested families are presented in **Figure 1**.

Considering the survival rate, growth stage, and pedigree information, three resistant families (VR4013, VR3837, and VR4027) and three susceptible families (VS3868, VS3879, and VS3880) were selected for transcriptome sequencing. The survival rates of three resistant families VR4013 (1.96 ± 0.21 g), VR3837 (2.14 ± 0.33 g), and VR4027 (2.21 ± 0.26 g) were 87%, 96%, and 78%, and those of three susceptible families VS3868 (1.87 ± 0.24 g), VS3879 (2.09 ± 0.28 g), VS3880 (2.19 ± 0.30 g) were 4%, 9%, and 13%, respectively (**Figure 1**). Based on their pedigree information, the families VR4013 and VS3868 were genetically related; therefore, the DEGs were analyzed between VR4013 and VS3868 in order to avoid the effects of genetic difference as far as possible. According to information of the growth stage and pedigree information, DEGs were analyzed between VR3837 and VS3879, and VR4027 and VS3880, respectively.

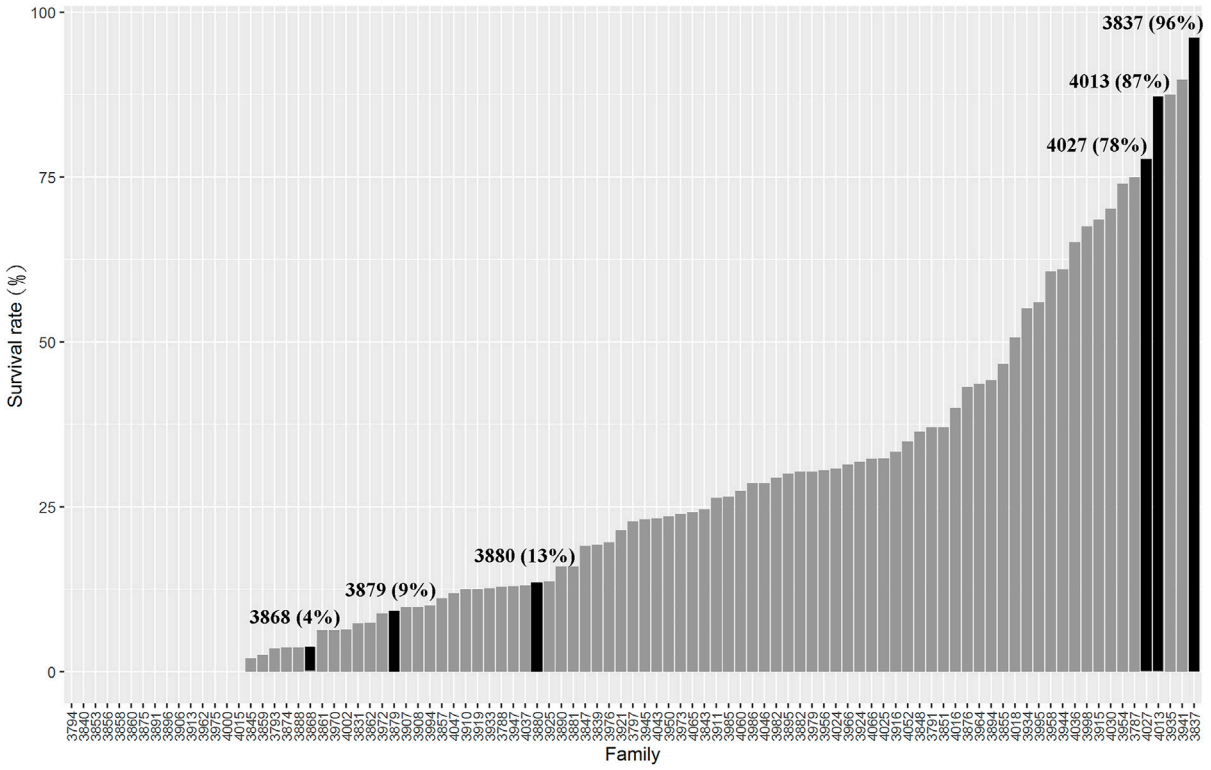

FIGURE 1 | Survival rate of different families challenged by VP_{AHPND}.

For each family, the cephalothoraxes of nine individuals were collected, and three individuals were mixed together as one sample, so each family contained three samples. All cephalothorax samples of the six families were rapidly frozen in liquid nitrogen and stored at −80°C for further transcriptome sequencing.

Total RNA Extraction, Library Construction, and Sequencing

Total RNA of the cephalothorax was extracted using RNAiso Plus (Takara, Japan) according to the manufacturer's instructions. RNA purity and concentration were evaluated by electrophoresis on 1% agarose gel and quantified by NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Subsequently, mRNA was enriched by Oligo (dT) beads. Then the enriched mRNA was fragmented into short fragments using fragmentation buffer and reverse-transcribed into the first-strand cDNA with random primers. Second-strand cDNA was synthesized by DNA polymerase I, RNase H, dNTP, and buffer. Then the cDNA fragments were purified with QiaQuick PCR extraction kit, end repaired, poly (A) added, and ligated to Illumina sequencing adapters. The suitable fragments were selected by agarose gel electrophoresis. After PCR amplification, the libraries were sequenced using Illumina HiSeqTM 2,500 by Genedenovo Biotechnology Co., Ltd (Guangzhou, China).

Bioinformatics Analysis

Clean reads were obtained by removing reads containing adapters or more than 10% of unknown nucleotides (N), low-quality reads, and rRNA from the raw data. Then the reads of each sample were mapped to reference genome (Zhang X. et al., 2019) by TopHat2 (version 2.0.3.12) (Kim et al., 2013). The reconstruction of transcripts was carried out with software Cufflinks (Trapnell et al., 2012). All reconstructed transcripts were aligned to reference genome, and novel genes were aligned to databases including National Center for Biotechnology Information (NCBI) nonredundant (Nr) database, Swiss-Prot database, and Kyoto Encyclopedia of Genes and Genomes (KEGG) database to obtain protein functional annotation. DEGs between susceptible and resistant libraries were analyzed using the edgeR package (http://www.rproject.org/), with a fold change ≥2 and a false discovery rate (FDR) <0.05. DEGs were then subjected to enrichment analysis of Gene Ontology (GO) functions and KEGG pathways, which were performed using the OmicShare tools (http://www.omicshare.com/tools).

Gene Set Enrichment Analysis

As a complement to the differential expression analyses, gene set enrichment analysis (GSEA) for all 27,331 expressed genes was performed by GSEA software v4.10 (https://www.gsea-msigdb.org/gsea/downloads.jsp) (Subramanian et al., 2005). The submitted gene list was ranked by gene expression value using Signal2Noise method. GSEA was performed with default

TABLE 1 | Primer sequences and annealing temperature used for RT-qPCR.

Gene name	Forward primer (5′–3′)	Reverse primer (5′ to 3′)	Product size (bp)	Tm (°C)
ncbi_113822350	ATTCCCTACAGTGACGACTA	GCCAAAAGATTCTCTCATGC	132	51
ncbi_113828004	GTTCCAAACCCTACTTGTCT	CTATGTCCAAAACGGAATGC	106	51
XLOC_016514	AGATGCTTCCCTGGATCAACC	TGGACTCTCCATTCCGATGTTC	127	57
XLOC_016534	TCGCCATGAAGAACTGGTCA	GCAAATTGAAGGCGTCAGCA	96	56
ncbi_113824827	TTGCGAGACAGACCAACCAG	CAGGTGCAATCTTCATCGCC	133	55
ncbi_113816316	TTCCTCCCGCAAGACAAG	GAGGGAGGGTTGGGTTTT	150	55
ncbi_113816839	CTTTCGGGAGGGAGCGTAT	ACGGGAATAGTCCATCCAAGT	162	56
ncbi_113810874	ACCCGCTGTCCGCTCTACCA	TGTCCCAGCCGCAGCTCAAC	118	64
ncbi_113805286	GGGCAACTTACGGCTTCT	TTCGTGCCAATGGGTTTC	131	54
ncbi_113826200	ATTGCAGCACCGTCTCCT	TCCCTCAGGCAGACTTCG	91	57
XLOC_026751	TCTTGTGCCTCGCTGTGG	GGTGATGTGCGTGATCTTCTT	149	57
ncbi_113826199	CTCACCGCTGCGAGGATT	TCCCTCAGGCAGACTTCG	106	58
XLOC_023290	TCTGCTGGTGATGATGGT	GTCATCGGGAGAACAACT	142	52
ncbi_113808761	CCGCAATGCTGTAGAAGGAC	CGGCGGTCAGAGTGGAGAT	149	58
ncbi_113802520	CTTCTTGCCGTGTTTGCC	ACGATGCCGTCTCCTGTCT	164	57
ncbi_113815780	ACTCATAACCCACCGCCACT	TCGTCAGGGACCCAGCAA	154	59
ncbi_1,13820830	AAGCCGAACTTGGAGGACC	CGGATGAACTTACCGAAACG	110	56
XLOC_016349	CATCAAGCCCAAACCACC	TCTTCTCCAGCCAGCCACT	104	58
XLOC_016348	ATTGGACGCAAGGAGTATGG	CCTGGGCTGGTTGATGAG	146	55
ncbi_1,13817858	GAGGATGGGCTGAAATGTG	GTCCAGCAACTCTGAAGTATGA	155	54
ncbi_113825958	GGAACAGCAGACGGGAGTG	CAACGAAGCATTGGTGGC	93	57
ncbi_113828431	CCGTCACCAACACCCATAA	AGCAGCCACCCAAGGAAA	150	56
ncbi_113807930	GCTCGTCACCACAACCAT	CGAAGATGGGAGGCAGGT	145	56
ncbi_113816695	GAGCACCTCGCTTTCTGTTT	CATGACTTGGGTTCAGGTTTA	107	54
ncbi_113804592	TCACGGAGTGCCGCTACGAT	TCCCTGTTGCGGATGTCCTG	187	60
ncbi_113813557	CTGCCAGTGGAACACGCTAT	GCGGTGCTAGGAACGTAACTAA	186	57
ncbi_1,13817635	ATGCTGACAAGGCGAATA	AAGAGTCAGACCCGCAAG	159	52
ncbi_113830625	CGTGAATCGCAGTCCCTA	GTGGTCGCTTCCTCTTCC	100	55
ncbi_113807689	GGCAGCCGCATCTTCATC	AGGGCGAAGCGGCGGTTGTT	176	60
ncbi_1,13819349	GTTCCATACCGCCGTTACCA	CGAGCAATTTCGCTTACAACACTA	119	55
ncbi_113806536	GCACTTCCAAAGCCAACGA	GATCTCCTCGGAGTTGTAGCG	119	57
ncbi_113802817	CGTCGCTGGGCACAAGTA	AGCCGAAGTGTCCCGTTA	167	57
ncbi_1,13817262	AATGAGGCGGAGGAGCAG	CCTTCCAGGTGGCAGACAG	92	58
ncbi_113821874	TAAGAAGGTCCAGAGGCG	AACCCACAAGGCCATACA	125	55
ncbi_113810465	CGCTGGTGGGTGTCGTGAT	CCGCTTGGCTGCTGAGAT	117	55
ncbi_113811111	CCGAGGTCAACTACGAGG	ACGGGACTTGGTGGCTGGT	106	55
ncbi_113816327	AAACCCAACCCTCCCTCT	TCCTCCGTCTCCAACACC	136	54
ncbi_113821801	AAGGGCGTGGAAGGAATG	CTTCATCTCCTCCTTCTCCTTC	187	56
18S	TATACGCTAGTGGAGCTGGAA	GGGGAGGTAGTGACGAAAAAT	136	56

algorithm as 1,000 permutations. GO gene sets (10,192 gene sets) and KEGG subset of canonical pathways (186 gene sets) were used as enrichment input, which were from Molecular Signatures database (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp). Nominal p-value < 0.05 and FDR < 0.25 were considered as statistically significant.

Evaluation on the Transcriptome Results by Real-Time Quantitative PCR

Six genes were selected from each comparison group to evaluate the transcriptome sequencing result by RT-qPCR. About 1 mg of total RNA was used to synthesize cDNA by PrimeScript™ RT reagent Kit with gDNA Eraser kit (Takara, Japan) according to the manufacturer's instructions. Gene-specific primers (**Table 1**) were designed using Primer 5, where 18S rRNA was used as the reference gene for RT-qPCR analysis. RT-qPCR was conducted with the following conditions: denaturation at 94°C for 2 min, 40 cycles of 94°C for 30 s, annealing temperature for 20 s, and 72°C for 30 s. Each sample had four technical replicates. The specificity of the primer set was checked by melting curve analysis. The relative expression level was calculated with $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Validation of Candidate Genes in Descendant Families

In order to validate the identified DEGs, the descendant families of VS3868 and VR4013 were collected, and the DEGs were further validated in the descendant families by RT-qPCR. Family 4419 was produced by full-sib mating of family VS3868, and family 4253 was produced by full-sib mating of family VR4013. After being challenged by VP_{AHPND}, the survival rate of family 4253 was over 3.5 times higher than that of family

TABLE 2 | Summary of transcriptome sequencing and assembly of the transcriptome from *Litopenaeus vannamei*.

Sample	Raw reads	Clean reads	Percentage remained (%)	Gene number
VR4013-1	43,410,372	42,125,030	97.04	20,485
VR4013-2	57,672,500	55,606,214	96.42	20,677
VR4013-3	51,479,718	49,787,374	96.71	21,014
VR3837-1	38,338,506	36,947,962	96.37	19,828
VR3837-2	37,704,754	36,467,784	96.72	19,592
VR3837-3	47,019,886	45,527,220	96.83	20,211
VR4027-1	56,229,492	54,118,212	96.25	21,034
VR4027-2	39,363,752	37,908,938	96.30	20,832
VR4027-3	44,978,148	43,327,096	96.33	20,310
VS3868-1	42,426,736	41,120,350	96.92	20,912
VS3868-2	42,975,552	41,754,004	97.16	20,621
VS3868-3	41,490,144	40,033,320	96.49	20,573
VS3880-1	51,220,018	49,705,886	97.04	20,157
VS3880-2	45,819,944	44,689,240	97.53	20,289
VS3880-3	42,021,826	40,653,124	96.74	20,005
VS3879-1	49,914,612	48,181,284	96.53	20,111
VS3879-2	43,730,322	42,164,296	96.42	19,971
VS3879-3	47,364,768	45,863,282	96.83	20,366

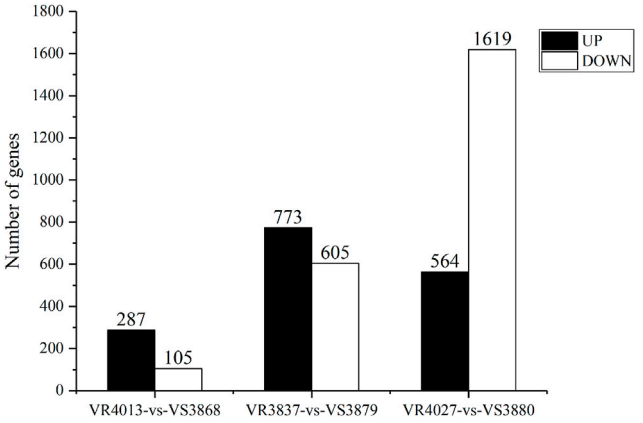

FIGURE 2 | The amount of differentially expressed genes (DEGs) between resistant and susceptible families. UP represents highly expressed genes in susceptible families. DOWN represents highly expressed genes in resistant families.

4419 (60 vs. 17%). A total of 32 DEGs including the 16 DEGs shared in three pairwise comparisons and other 16 DEGs shared in two pairwise comparisons were selected, and they were verified in family 4419 and 4253. The primers designed by Primer 5 are listed in **Table 1**. The RT-qPCR was performed as described in *Evaluation on the Transcriptome Results by Real-Time Quantitative PCR*. A statistically significant difference was indicated with * ($p < 0.05$) and ** ($p < 0.01$) determined by *t*-test (SPSS Inc., Armonk, NY, USA).

RESULTS

Transcriptome Sequencing Data

An overview of sequencing and assembly of the *L. vannamei* transcriptome is shown in **Table 2**. A total of 823,161,050 raw reads were obtained, in which 416,197,128 reads are for the resistant families and 406,963,922 reads for the susceptible families. The raw sequencing data were uploaded to the NCBI with the accession numbers SRR15533118–SRR15533135. After low-quality reads were filtered out, a total of 96.55% and 96.85% reads were retained for the resistant and susceptible groups, respectively. After the reads were mapped to the reference genome, a total of 27,331 annotated genes were obtained, of which 2,344 (9.86%) were newly annotated genes.

Identification of Differentially Expressed Genes Between Resistant and Susceptible Families

To identify DEGs involved in VP_{AHPND} resistance, we used FPKM value for comparing the expression levels between the resistant families and susceptible families. A total of 392 unigenes showed differential expression patterns between families VR4013 and VS3868, in which 287 unigenes were highly expressed in VS3868 and 105 unigenes were highly expressed in VR4013 (**Figure 2**). A total of 1,378 unigenes were differentially expressed between VR3837 and VS3879, including 773 unigenes highly expressed in VS3879 and 605 unigenes highly expressed in VR3837. A total of 2,183 unigenes were differentially expressed between VR4027 and VS3880, including 564 unigenes highly expressed in VS3880 and 1,619 unigenes highly expressed in VR4027. Six DEGs selected from each comparison group were validated by RT-qPCR. The results showed that all of them were consistent with the transcriptome data (**Supplementary Figure S1**).

In order to analyze the function of DEGs between resistant and susceptible families, GO functional enrichment analysis was performed. Interestingly, we found that GO terms of DEGs were very similar among three pairs of families, VR4013-vs.-VS3868, VR3837-vs.-VS3879, and VR4027-vs.-VS3880, which is shown in **Figure 3A**. In the biological process category, single-

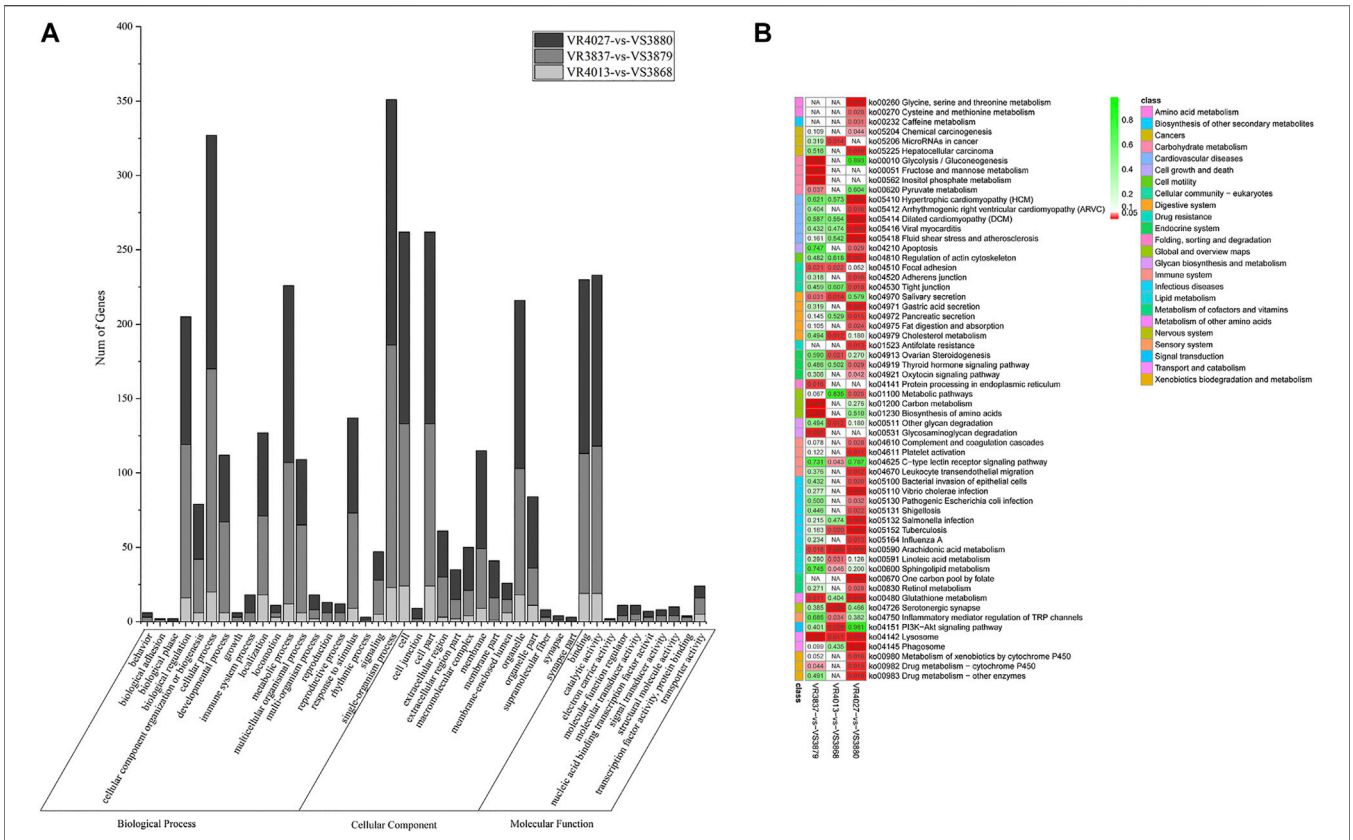

FIGURE 3 | Enrichment analysis of differentially expressed genes (DEGs). **(A)** Gene Ontology (GO) terms distribution for the DEGs. Light gray represents VR4013-vs.-VR3868. Gray represents VR3837-vs.-VR3879. Dark gray represents VR4027-vs.-VR3880. The *x*-axis indicates the name of GO subcategories. The *y*-axis represents the number of genes. **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched about DEGs in VR4013-vs.-VR3868, VR3837-vs.-VR3879, and VR4027-vs.-VR3880.

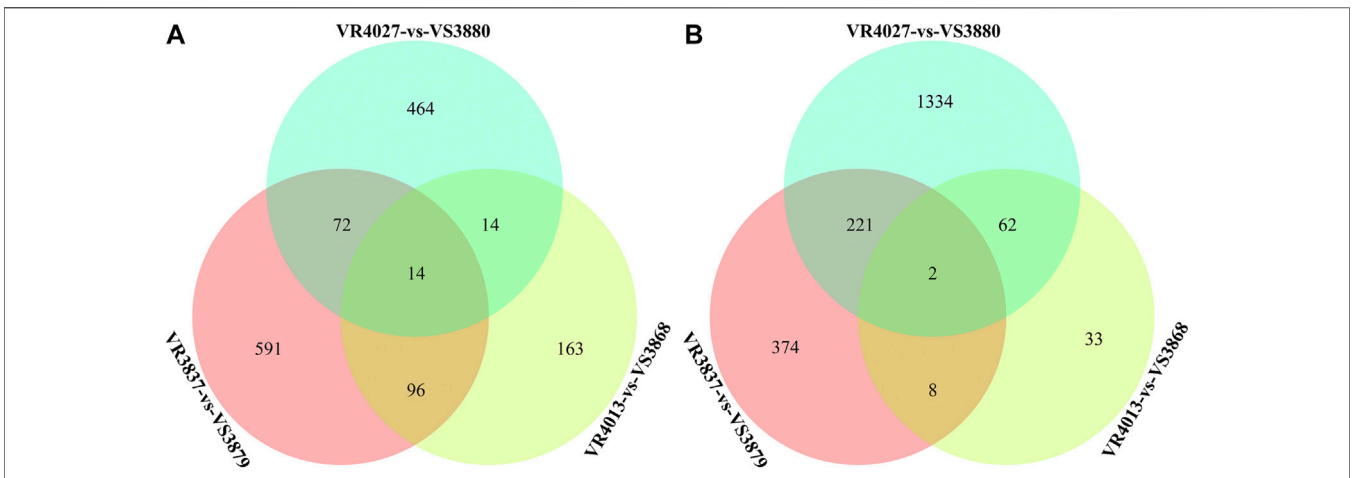

FIGURE 4 | Venn diagrams for three pairwise comparisons of resistant and susceptible families. **(A)** Differentially expressed genes (DEGs) with high expression in susceptible families. **(B)** DEGs with high expression in resistant families.

organism process was the most enriched subclasses, followed by cellular process. In the cellular component category, cell and cell part were the two most enriched subclasses. While in the molecular function category, catalytic activity and binding were the two most enriched subclasses. KEGG pathway enrichment analysis of three pairwise comparisons is

TABLE 3 | Gene name, gene annotation, and the verification in hepatopancreas (Hep) and stomach (St) of 16 DEGs that were shared in three pairwise comparisons.

Gene name	Gene annotation	Verification (*p*-value)	
		Hep	St
ncbi_113805286	NA[a]		√(0.001)
ncbi_113810874	Trophoblast glycoprotein		√(0.024)
ncbi_113826200	NA	√(0.025)	
XLOC_016514	NA		√(0.004)
XLOC_016534	NA		
ncbi_113822350	SE-cephalotoxin-like	√(0.004)	√(0.01)
ncbi_113824827	Prolow-density lipoprotein receptor-related protein 1-like	√(0.018)	√(0.013)
ncbi_113828004	Peroxidasin-like protein		√(0.02)
ncbi_113819349	Phosphoenolpyruvate carboxykinase, cytosolic		
ncbi_113807689	Cytochrome P450		
XLOC_026751	NA	√(0.006)	√(0.014)
ncbi_113826199	Single insulin-like growth factor-binding domain protein-2	√(0.003)	√(0.009)
XLOC_023290	NA	√(0.012)	
ncbi_113825958	Serpin B6-like isoform X2		
ncbi_113816327	NA		
ncbi_113816316	NA	√(0.023)	

Note. The threshold for significance is p*-value < 0.05.*
DEGs, differentially expressed genes.
[a]*NA indicates that the function of gene is unknown.*

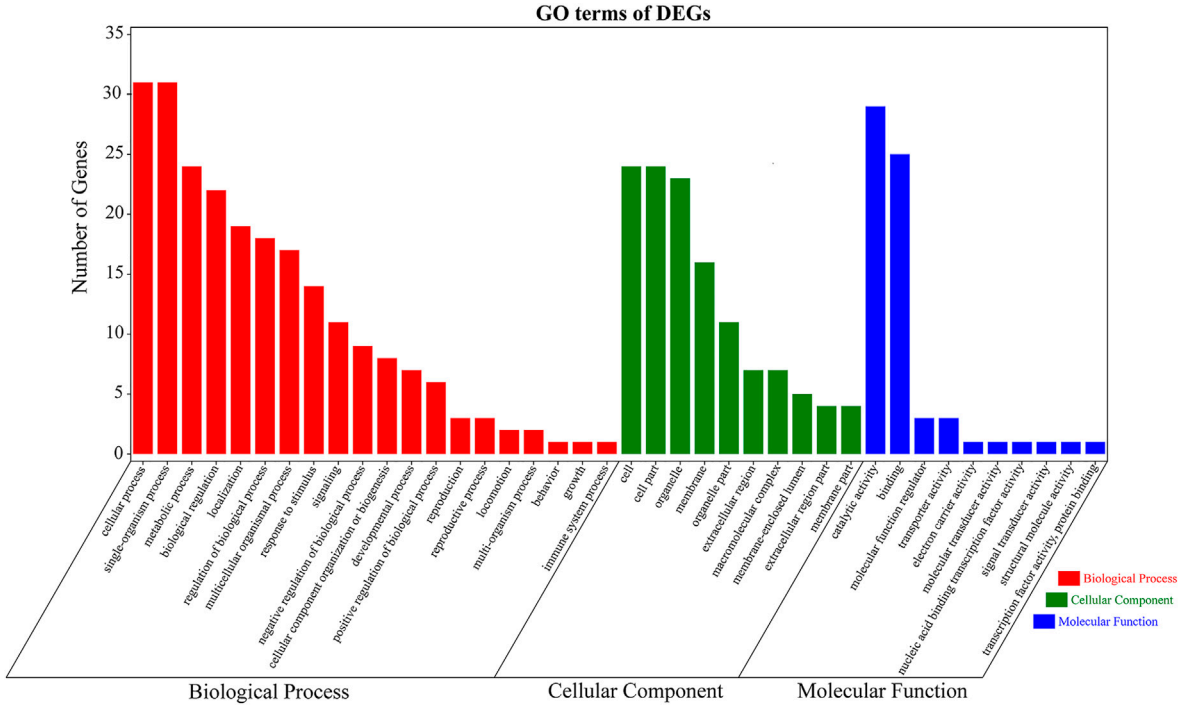

FIGURE 5 | Gene Ontology (GO) terms of 489 differentially expressed genes (DEGs) for the transcriptomes of *Litopenaeus vannamei*. The *x*-axis indicates the name of GO subcategories. The *y*-axis indicates the number of genes. Red indicates biological process. Green indicates cellular component. Blue displays molecular function.

presented in **Figure 3B**. In VR4013-vs.-VS3868, "Serotonergic synapse," "Arachidonic acid metabolism," and "PI3K–Akt signaling pathway" were the three most enriched pathways, and "Arachidonic acid metabolism" was also included in top 20 of enrichment pathways in VR3837-vs.-VS3879 and VR4027-vs.-VS3880 ($p < 0.05$).

Differentially Expressed Genes Shared in More Than Two Pairwise Comparisons

Venn diagrams for DEGs of the three pairwise comparisons are shown in **Figure 4**. A total of 489 DEGs were shared in at least two comparisons, including 196 DEGs highly expressed in the susceptible families and 293 DEGs highly expressed in resistant

TABLE 4 | Myosin- and energy metabolism-related DEGs with high expression in VP_{AHPND}-resistant *Litopenaeus vannamei* families.

Gene ID	Description	Species
ncbi_113815780	Myosin-16-like	*Lepisosteus oculatus*
ncbi_113820830	Myosin-16 isoform X2	*L. oculatus*
XLOC_031566	Myosin heavy chain, muscle-like isoform X7	*Hyalella azteca*
ncbi_113816531	Myosin-7	*Chelonia mydas*
XLOC_016349	Myosin heavy chain type 2	*L. vannamei*
XLOC_016348	Myosin heavy chain type 2	*Penaeus monodon*
XLOC_016355	Myosin heavy chain type b	*Marsupenaeus japonicus*
XLOC_015656	Myosin heavy chain type b	*M. japonicus*
ncbi_113802000	Myosin heavy chain, partial	*Rana catesbeiana*
XLOC_009141	Actin 2	*Nilaparvata lugens*
ncbi_113815891	Tryptase-like	*Lates calcarifer*
ncbi_113822347	Tryptase-2-like	*Scleropages formosus*
ncbi_113813760	Trypsin-like	*Sinocyclocheilus rhinocerous*
ncbi_113809217	Trypsin-2-like	*S. rhinocerous*
ncbi_113814573	Chymotrypsinogen A	*Ovis aries musimon*
ncbi_113806996	Serine protease 33-like	*Python bivittatus*
ncbi_113800311	Serine protease 27-like	*L. calcarifer*
XLOC_005102	Aminopeptidase N-like	*H. azteca*
ncbi_113815289	Aminopeptidase N	*Ornithorhynchus anatinus*
ncbi_113815282	Aminopeptidase N	*Octodon degus*
ncbi_113815313	Aminopeptidase N	*Ochotona princeps*
ncbi_113815275	Aminopeptidase Ey-like	*Xenopus laevis*
ncbi_113811393	Aminopeptidase Ey-like	*X. laevis*
ncbi_113804819	Aminopeptidase Ey-like	*X. laevis*
ncbi_113804804	Aminopeptidase N, partial	*Chlorocebus sabaeus*
XLOC_009027	Cathepsin I, partial	*L. vannamei*
ncbi_113807506	Pancreatic lipase-related protein 2-like	*Chaetura pelagica*
ncbi_113821709	Phospholipase A2	*Pagrus major*
ncbi_113808761	Alkaline phosphatase, tissue-nonspecific isozyme-like	*S. rhinocerous*
ncbi_1,13820123	Glyceraldehyde 3-phosphate dehydrogenase	*Esox lucius*
ncbi_113802551	Triosephosphate isomerase A	*Astyanax mexicanus*

Note. DEGs, differentially expressed genes.

families (**Supplementary Table S1**). A total of 16 DEGs were shared in three pairwise comparisons, among which the trophoblast glycoprotein, SE-cephalotoxin-like, peroxidasin-like protein, phosphoenolpyruvate carboxykinase, cytochrome P450, and serpin B6-like isoform X2 genes were identified (**Table 3**). In order to reduce the false positives, we considered these 489 DEGs as the candidate genes associated with the resistance of shrimp against to VP_{AHPND}.

These 489 DEGs were involved in various GO classifications. For the biological process-related genes, most were involved in "single-organism process," "cellular process," and "metabolic process." Most of the cellular component-related genes were associated with "cell," "cell part," and "organelle." And "catalytic activity" and "binding" in the molecular function ontology were the major enriched terms (**Figure 5**). The results were consistent with GO functional enrichment analysis in **Figure 3A**.

Differentially Expressed Genes with High Expression in Resistant Families

There were 293 DEGs highly expressed in resistant families. A total of nine genes annotated as myosin were highly expressed, and the other genes were involved in energy metabolism such as trypsin, chymotrypsinogen A (ChyA), pancreatic lipase (PL), serine protease (SP), aminopeptidase, and phospholipase (**Table 4**). There were two genes encoding glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and triosephosphate isomerase (TPI), which were involved in the glycolysis pathway (Eanes, 2011).

Differentially Expressed Genes with High Expression in Susceptible Families

A total of 196 DEGs highly expressed in the susceptible families were discovered. According to annotation and function, some immune-related genes were observed to be upregulated in the susceptible families (**Table 5**). A total of five genes related to the prophenoloxidase (proPO) system, including C-type lectin, SP, and SP inhibitors (serpin and α-2-macroglobulin) showed high expression in the susceptible families. Several genes are related to metabolic process, including DBH-like monooxygenase protein 1, cytochrome b5, and cytochrome P450. Moreover, six genes annotated as β-arrestin were also highly expressed in the VP_{AHPND}-susceptible families.

Gene Set Enrichment Analysis

By GSEA of GO gene set, there were a total of 265 significantly enriched gene sets between resistant and susceptible groups (nominal p-value < 0.05 and FDR < 0.25). The top 10 gene sets enriched in resistant families were all involved in energy metabolism process, while most gene sets enriched in the

TABLE 5 | Immunity-related DEGs with high expression in VP_{AHPND}-susceptible *Litopenaeus vannamei* families.

Gene ID	Description	Species
ncbi_113817858	C-type lectin domain family 4 member E-like isoform X3	*Lates calcarifer*
ncbi_113825958	Serpin B6-like isoform X2	*Sarcophilus harrisii*
ncbi_113828431	Serine protease 42-like	*Chrysochloris asiatica*
ncbi_113808061	α-2-Macroglobulin-like protein 1	*Pseudopodoces humilis*
ncbi_113808062	α-2-Macroglobulin-like protein 1	*Coturnix japonica*
ncbi_113816695	DBH-like monooxygenase protein 1	*Otolemur garnettii*
ncbi_113804592	DBH-like monooxygenase protein 1	*O. garnettii*
ncbi_113813557	Cytochrome b5-like	*Salmo salar*
ncbi_113807689	Cytochrome P450	*Danio rerio*
ncbi_113829525	β-Arrestin-2 isoform X2	*Heterocephalus glaber*
ncbi_113826122	β-Arrestin-2	*Pelodiscus sinensis*
ncbi_113815019	β-Arrestin-2	*Lepisosteus oculatus*
ncbi_113822594	β-Arrestin-1-like, partial	*Takifugu rubripes*
ncbi_113804603	β-Arrestin-1, partial	*Kryptolebias marmoratus*
ncbi_113807566	β-Arrestin-1	*Gekko japonicus*

Note. DEGs, differentially expressed genes.

TABLE 6 | GSEA of GO gene sets.

EP	Gene set	Size	NES	Nom *p*-val	FDR
R	Mitochondrial_protein_complex	78	2.70	0	0
R	Inner_mitochondrial_membrane_protein_complex	36	2.64	0	0
R	Cellular_respiration	64	2.61	0	0
R	Oxidative_phosphorylation	44	2.58	0	0
R	Mitochondrial_respiratory_chain_complex_assembly	33	2.57	0	0
R	Respiratory_electron_transport_chain	40	2.54	0	0
R	Respiratory_chain_complex	27	2.53	0	0
R	ATP_synthesis_coupled_electron_transport	35	2.52	0	0
R	NADH_dehydrogenase_complex_assembly	26	2.49	0	0
R	Respirasome	35	2.44	0	0
S	G_Protein_coupled_receptor_activity	52	−2.16	0	0.119
S	G_Protein_coupled_receptor_signaling_pathway	115	−2.11	0	0.130
S	Tight_junction	18	−2.10	0	0.101
S	Positive_regulation_of_blood_circulation	8	−2.10	0	0.078
S	Solute_sodium_symporter_activity	32	−2.10	0	0.066
S	Neurotransmitter_binding	17	−2.10	0	0.055
S	Neuromuscular_process_controlling_balance	10	−2.09	0	0.048
S	Apical_junction_complex	18	−2.05	0.003	0.072
S	Serine_type_endopeptidase_inhibitor_activity	17	−2.05	0.003	0.069
S	Pattern_recognition_receptor_signaling_pathway	27	−2.02	0	0.085

Note. EP, enrichment in phenotype, gene sets enriched in nine resistant (R) samples or nine susceptible (S) samples; Size, number of genes in the gene set; NES, normalized enrichment score; NOM p-value, nominal p-value, the statistical significance of the enrichment score; FDR, false discovery rate; GSEA, Gene set enrichment analysis; GO, Gene Ontology.

susceptible families were related to signal transduction and immunity, such as G protein-coupled receptor activity, G protein-coupled receptor signaling pathway, and pattern recognition receptor signaling pathway (**Table 6**). **Figure 6** shows GSEA of KEGG subset of canonical pathways. The most significantly enriched pathway in resistant families was oxidative phosphorylation (**Figure 6A**), which was also related to energy metabolism, supporting the results of GO gene sets. The most significantly enriched pathway in the susceptible families was JAK/STAT signaling pathway (**Figure 6B**), including protein inhibitor of activated STAT, cytokine-inducible SH2-containing protein, and tyrosine-protein phosphatase non-receptor type 11 isoform X2.

Verification of Candidate Genes in the Descendant Families

A total of 32 genes were validated in descendant families. For the 16 DEGs shared by three pairwise comparisons, seven genes were verified successfully in the hepatopancreas, and eight genes were verified successfully in the stomach in their offspring (**Table 3**, $p < 0.05$). For the other 16 DEGs shared by two pairwise comparisons, three genes were successfully verified in hepatopancreas ($p < 0.05$) (**Figure 7A**), and six genes were verified successfully in the stomach ($p < 0.05$) (**Figure 7B**). In summary, a total of 19 genes were successfully verified in the hepatopancreas or stomach, including five genes verified successfully in two tissues.

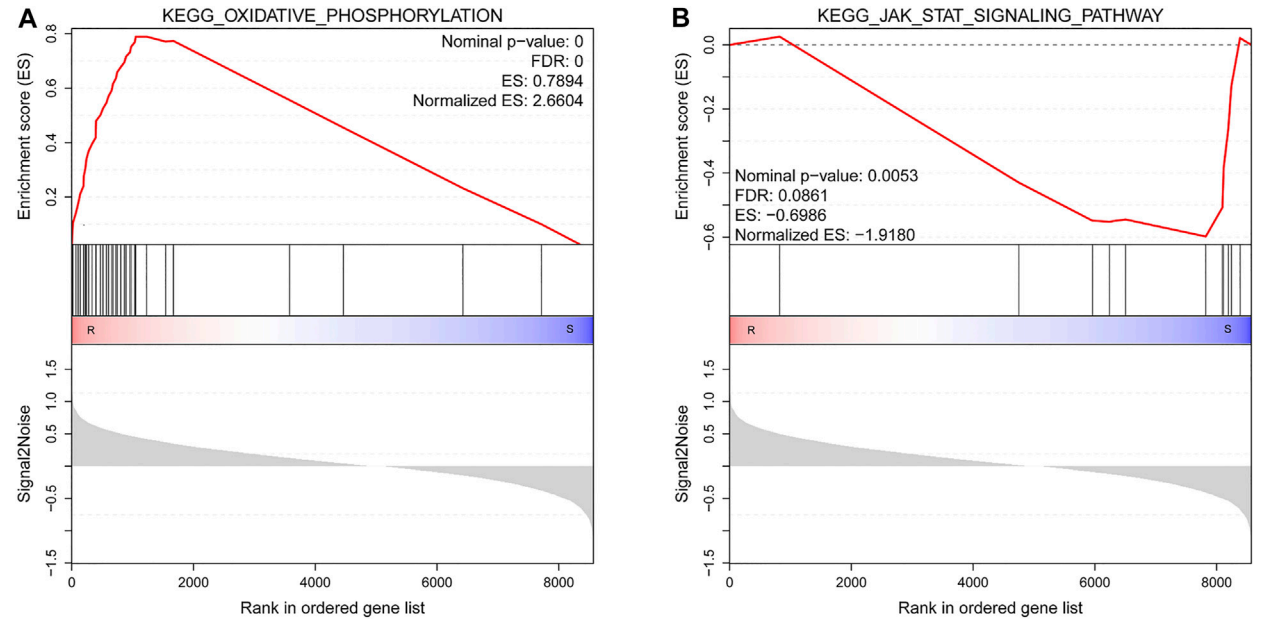

FIGURE 6 | Gene set enrichment analysis (GSEA) of Kyoto Encyclopedia of Genes and Genomes (KEGG) subset of canonical pathways. **(A)** The most significantly enriched pathway in resistant families. **(B)** The most significantly enriched pathway in susceptible families. R, resistant samples; S, susceptible samples. Nominal *p*-value, false discovery rate (FDR), enrichment score (ES), and normalized ES were determined by the GSEA software and were indicated within each enrichment plot.

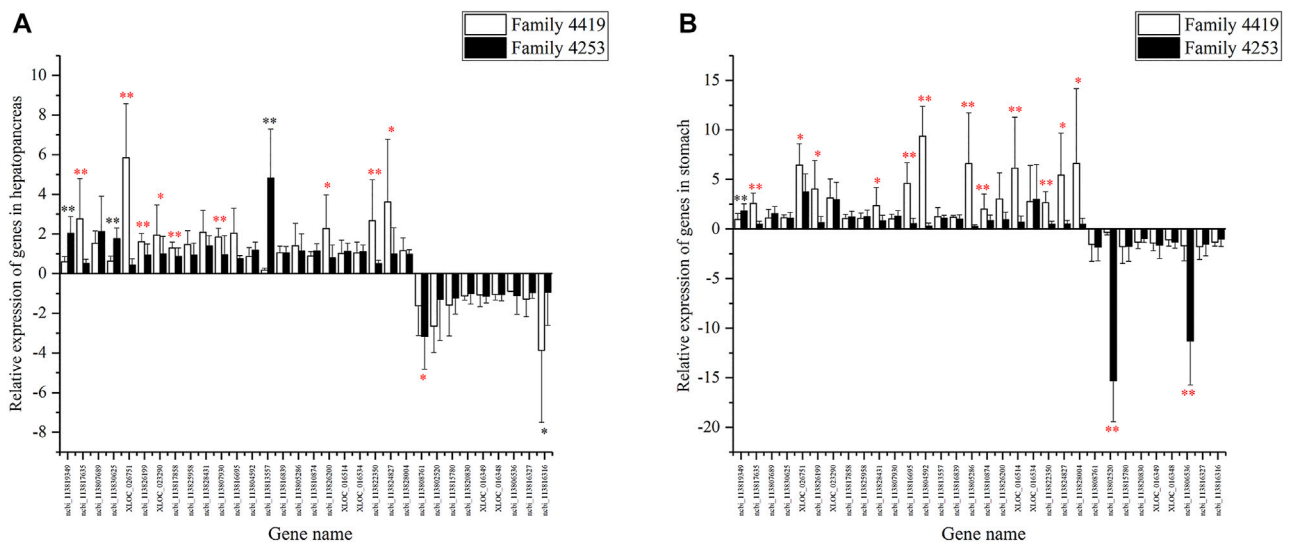

FIGURE 7 | The relative expression of 32 differentially expressed genes (DEGs) in hepatopancreas **(A)** and stomach **(B)** from resistant family 4253 and susceptible family 4019. These data are expressed as the mean ± SD relative to the reference gene (18S rRNA). A statistically significant difference is indicated with * ($p < 0.05$) and ** ($p < 0.01$). Successfully verified genes are marked in red. Unsuccessfully verified genes are marked in black or without any mark.

DISCUSSIONS

It is generally considered that the VP_{AHPND} is an opportunistic pathogen. It is pathogenic to cultured shrimp at high concentration or under environmental deterioration condition (Hong et al., 2016). The physiology and health state of shrimp are closely related to disease resistance. It was estimated that the heritability of the resistance against VP_{AHPND} in *L. vannamei* was near-to-moderate, which indicated that the resistance of shrimp against VP_{AHPND} was hereditable (Wang et al., 2019). In this

study, the resistant and susceptible families of *L. vannamei* were obtained through continuous challenge test, which provided reliable genetic material for analysis on the disease resistance of shrimp (Grimholt, et al., 2003; Wang et al., 2014). Understanding the gene expression character of disease-resistant families and the susceptible families could provide useful information for genetic dissection of disease resistance of shrimp.

Recently, an increasing number of transcriptome studies related to AHPND have been performed in *L. vannamei*, most of which focused on the genes involved in the immune response of shrimp during VP_{AHPND} infection (Qi et al., 2017; Maralit et al., 2018; Ng et al., 2018; Qin et al., 2018; Zheng et al., 2018). However, few works have focused on the correlation between the gene expression level and resistance phenotype of shrimp. In this study, we performed RNA-seq analyses to investigate the comparative expression profiles between resistant and susceptible families of shrimp. Gene expression profiles can underlie complex phenotype variations (Crawford and Powers, 1992). The gene expression variation is influenced by various genetic and environmental factors (Leder et al., 2014). Many studies proved that genetic factors influence gene expression in humans (Cheung and Spielman, 2002), mice (Sandberg et al., 2000), *Drosophila* (Jin et al., 2001), and yeast (Brem et al., 2002). To our knowledge, this is the first report about the comparison of the basal mRNA expression profiles from the omics level of AHPND-resistant and AHPND-susceptible families of *L. vannamei*. The gene expression levels influencing the phenotype variations could also be considered as molecular markers and could be used for genetic selection (Gilad et al., 2008). For aquaculture species, genes related to phenotype variations have been reported. The expression of BIRC7 was correlated with the survival in *Vibrio* challenge tests in clam *M. petechialis*, and gene expression variation of BIRC7 gene was heritable, indicating the feasibility of selective breeding by reliable genetically based markers (Jiang et al., 2018; Jiang et al., 2019).

In the present study, we found that DEGs highly expressed in the susceptible families of shrimp were mainly related to the immunity, which were in agreement with previous research results. It was reported that the basal expression levels of immune-related genes BIRC7 and NFIL3 were higher in *Vibrio*-susceptible clams (Jiang et al., 2018). In this study, several genes in proPO activation system showed a higher expression level in the susceptible families. A previous report showed that PPAE2 presented a higher expression level in the susceptible line of *L. vannamei* after AHPND challenge in comparison with the resistant line (Mai et al., 2021). This system is important for fighting against bacteria pathogens in penaeid shrimp (Amparyup, et al., 2013), and the increased activity of proPO system against *Vibrio* has been reported in *Fenneropenaeus indicus* (Sarathi et al., 2007), *L. vannamei* (Boonchuen et al., 2021), and *Macrobrachium rosenbergii* (Rao et al., 2015). The C-type lectin was also upregulated in the susceptible families of shrimp; the C-type lectin plays an important role in phagocytosis, melanization, respiratory burst, and coagulation; and it can also activate the proPO system (Junkunlo et al., 2012). Interestingly, C-type lectin 1-like and Crustin-P had significant higher expression levels in the AHPND-susceptible line of *L. vannamei* than resistant line at 24 h post-infection (Mai et al., 2021). In *Litopenaeus stylirostris*, the basal expression level of antimicrobial peptide was significantly higher in *Vibrio*-susceptible shrimp line (Lorgeril et al., 2008). In addition, β-arrestin also played a vital role in the antibacterial immunity of shrimp (Jiang et al., 2013; Sun et al., 2016). The GSEA also illustrated that partial immune-related genes were upregulated in the susceptible families; the JAK/STAT signaling pathway, which was an important signal transduction pathway regulating the immune response in invertebrates, was significantly enriched in the susceptible families (Yu et al., 2017).

Besides, the genes in immunity showed differential expression patterns between susceptible and resistant families; another major part of the DEGs was related to myosin and metabolism. In this study, myosin and many other genes related to metabolism were more active in resistant families. Myosin and actin play a diverse role in a wide range of functions such as cytoskeleton, muscle contraction, and immune response (Marston 2018; Sitbon et al., 2019). It was already reported that myosin light chain was related to the phagocytosis against invading pathogens in *Penaeus japonicus*, and the transcription level of myosin in WSSV-resistant shrimp was nearly two times higher than that in the control shrimp (Han et al., 2010). After pathogen infection, myosin and actin were significantly upregulated in shrimp (Shi et al., 2018; Ren et al., 2019). As for metabolism, we found that enzymes like trypsin, ChyA, PL, and SP and glycolysis pathway including GAPDH and TPI were highly expressed in the resistant families (**Table 4**). Meanwhile, the top 10 gene sets enriched in resistant families in the GSEA were all involved in energy metabolism process. The finding was consistent with the previous report that SP and ChyB had a significantly higher expression level in resistant shrimp line during the AHPND infection (Mai et al., 2021). Several studies have indicated that metabolic processes such as lipid metabolism and carbohydrate metabolic process in shrimp were greatly affected during AHPND infection (Velazquez-Lizarraga et al., 2019; Kumar et al., 2021). Taken together, the activated myosin, actin, and energy metabolism might indicate that the shrimp were healthier, which led to higher resistance of shrimp to disease.

After validation, several genes showed the same expression pattern in the offspring of susceptible and resistant families. These genes are possible to be developed as biomarkers for disease resistance of shrimp. It was already reported that gene expression profile could be used as an indicator for disease resistance trait. For example, Bsr-d1 RNA level in susceptible rice strain was twofold higher than that in resistant rice strain, and it has been further proved that inhibiting the expression of bsr-d1 could increase the rice resistance against blast infection (Li et al., 2017). In *E. carinicauda*, eight immune-related genes were suggested as potential disease-resistant parameters for evaluating the physiological status and disease-resistant capability of shrimp when infected with VP_{AHPND} (Ge et al., 2018). In this study, the 19 genes successfully verified in their descendant families were expected to be developed as biomarkers of shrimp resistance against *Vibrio*. Therefore, apart from sib challenge testing and molecular marker-assisted breeding, the gene expression level of

these 19 genes could also be used as molecular markers for accelerating the breeding of disease-resistant varieties in *L. vannamei*.

In summary, this study integrated the VP_{AHPND}-resistance phenotype variation and gene expression profiles to identify the genes related to disease resistance of shrimp. A total of 489 DEGs were identified between the resistant and susceptible families, and they were considered to be associated with the ability of VP_{AHPND} resistance in *L. vannamei*. Gene annotation and enrichment analysis revealed that the immune response and energy metabolism could influence resistance of shrimp against VP_{AHPND}. The obtained data provide a fundamental basis for clarifying the genetic mechanism of resistance to bacterial pathogen, and the identified disease resistance genes of shrimp could accelerate the genetic breeding in shrimp aquaculture.

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**.

AUTHOR CONTRIBUTIONS

QZ, YY, and FL contributed to the conception and design of the study. ZL performed the statistical analysis. QZ wrote the first draft of the article. QZ, YY, JX, and FL wrote sections of the article. All authors contributed to article revision and read and approved the submitted version.

FUNDING

This work is supported by the key program of the National Natural Science Foundation of China (31830100), National Key R&D Program of China (2018YFD0901301, 2018YFD0900103), and China Agriculture Research System-48.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.772442/full#supplementary-material

REFERENCES

Amparyup, P., Charoensapsri, W., and Tassanakajon, A. (2013). Prophenoloxidase System and its Role in Shrimp Immune Responses against Major Pathogens. *Fish Shellfish Immunol.* 34 (4), 990–1001. doi:10.1016/j.fsi.2012.08.019

Argue, B. J., Arce, S. M., Lotz, J. M., and Moss, S. M. (2002). Selective Breeding of Pacific White Shrimp (*Litopenaeus vannamei*) for Growth and Resistance to Taura Syndrome Virus. *Aquaculture* 204, 447–460. doi:10.1016/s0044-8486(01)00830-4

Arora, S., Steuernagel, B., Gaurav, K., Chandramohan, S., Long, Y., Matny, O., et al. (2019). Resistance Gene Cloning from a Wild Crop Relative by Sequence Capture and Association Genetics. *Nat. Biotechnol.* 37 (2), 139–143. doi:10.1038/s41587-018-0007-9

Boonchuen, P., Jaree, P., Somboonviwat, K., and Somboonwiwat, K. (2021). Regulation of Shrimp Prophenoloxidase Activating System by Lva-miR-4850 during Bacterial Infection. *Sci. Rep.* 11 (1), 3821. doi:10.1038/s41598-021-82881-2

Brem, R. B., Yvert, G., Clinton, R., and Kruglyak, L. (2002). Genetic Dissection of Transcriptional Regulation in Budding Yeast. *Science* 296 (5568), 752–755. doi:10.1126/science.1069516

Cheung, V. G., and Spielman, R. S. (2002). The Genetics of Variation in Gene Expression. *Nat. Genet.* 32, 522–525. doi:10.1038/ng1036

Crawford, D. L., and Powers, D. A. (1992). Evolutionary Adaptation to Different thermal Environments via Transcriptional Regulation. *Mol. Biol. Evol.* 9, 806–813. doi:10.1093/oxfordjournals.molbev.a040762

de Lorgeril, J., Gueguen, Y., Goarant, C., Goyard, E., Mugnier, C., Fievet, J., et al. (2008). A Relationship between Antimicrobial Peptide Gene Expression and Capacity of a Selected Shrimp Line to Survive a *Vibrio* Infection. *Mol. Immunol.* 45 (12), 3438–3445. doi:10.1016/j.molimm.2008.04.002

Eanes, W. F. (2011). Molecular Population Genetics and Selection in the Glycolytic Pathway. *J. Exp. Biol.* 214, 165–171. doi:10.1242/jeb.046458

Ge, Q., Li, J., Li, J., Wang, J., and Li, Z. (2018). Immune Response of *Exopalaemon carinicauda* Infected with an AHPND-Causing Strain of *Vibrio parahaemolyticus*. *Fish Shellfish Immunol.* 74, 223–234. doi:10.1016/j.fsi.2017.12.042

Gilad, Y., Rifkin, S. A., and Pritchard, J. K. (2008). Revealing the Architecture of Gene Regulation: the Promise of eQTL Studies. *Trends Genet.* 24 (8), 408–415. doi:10.1016/j.tig.2008.06.001

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., et al. (2003). MHC Polymorphism and Disease Resistance in Atlantic salmon (*Salmo salar*); Facing Pathogens with Single Expressed Major Histocompatibility Class I and Class II Loci. *Immunogenetics* 55 (4), 210–219. doi:10.1007/s00251-003-0567-8

Han, F., Wang, Z., and Wang, X. (2010). Characterization of Myosin Light Chain in Shrimp Hemocytic Phagocytosis☆. *Fish Shellfish Immunol.* 29 (5), 875–883. doi:10.1016/j.fsi.2010.07.030

Hong, X. P., Xu, D., Zhuo, Y., Liu, H. Q., and Lu, L. Q. (2016). Identification and Pathogenicity of *Vibrio Parahaemolyticus* Isolates and Immune Responses of *Penaeus (Litopenaeus) vannamei* (Boone). *J. Fish. Dis.* 39 (9), 1085–1097. doi:10.1111/jfd.12441

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

Huang, Y., Yin, Z., Weng, S., He, J., and Li, S. (2012). Selective Breeding and Preliminary Commercial Performance of *Penaeus vannamei* for Resistance to white Spot Syndrome Virus (WSSV). *Aquaculture* 364-365, 111–117. doi:10.1016/j.aquaculture.2012.08.002

Jiang, D., Xie, T., Liang, J., and Noble, P. W. (2013). β-Arrestins in the Immune System. *Prog. Mol. Biol. Transl Sci.* 118, 359–393. doi:10.1016/b978-0-12-394440-5.00014-0

Jiang, F., Wang, H., Yue, X., Zhang, S., and Liu, B. (2018). Integrating the *Vibrio* -resistance Phenotype and Gene Expression Data for Discovery of Markers Used for Resistance Evaluation in the Clam *Meretrix petechialis*. *Aquaculture* 482, 130–136. doi:10.1016/j.aquaculture.2017.09.033

Jiang, F., Yue, X., Zhang, S., Yu, J., Wang, R., Liu, B., et al. (2019). Heritability of Resistance-Related Gene Expression Traits and Their Correlation with Body Size of Clam *Meretrix petechialis*. *J. Ocean. Limnol.* 38 (2), 571–578. doi:10.1007/s00343-019-8326-3

Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., and Gibson, G. (2001). The Contributions of Sex, Genotype and Age to Transcriptional Variance in *Drosophila melanogaster*. *Nat. Genet.* 29 (4), 389–395. doi:10.1038/ng766

Junkunlo, K., Prachumwat, A., Tangprasittipap, A., Senapin, S., Borwornpinyo, S., Flegel, T. W., et al. (2012). A Novel Lectin Domain-Containing Protein (LvCTLD) Associated with Response of the Whiteleg Shrimp *Penaeus (Litopenaeus) vannamei* to Yellow Head Virus (YHV). *Dev. Comp. Immunol.* 37, 334–341. doi:10.1016/j.dci.2011.12.010

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions. *Genome Biol.* 14 (4), R36. doi:10.1186/gb-2013-14-4-r36

Kumar, R., Tung, T.-C., Ng, T. H., Chang, C.-C., Chen, Y.-L., Chen, Y.-M., et al. (2021). Metabolic Alterations in Shrimp Stomach during Acute Hepatopancreatic Necrosis Disease and Effects of Taurocholate on *Vibrio parahaemolyticus*. *Front. Microbiol.* 12, 631468. doi:10.3389/fmicb.2021.631468

Leder, E. H., McCairns, R. J. S., Leinonen, T., Cano, J. M., Viitaniemi, H. M., Nikinmaa, M., et al. (2014). The Evolution and Adaptive Potential of Transcriptional Variation in Sticklebacks-Signatures of Selection and Widespread Heritability. *Mol. Biol. Evol.* 32, 674–689. doi:10.1093/molbev/msu328

Lee, C.-T., Chen, I.-T., Yang, Y.-T., Ko, T.-P., Huang, Y.-T., Huang, J.-Y., et al. (2015). The Opportunistic marine pathogen *Vibrio parahaemolyticus* becomes Virulent by Acquiring a Plasmid that Expresses a Deadly Toxin. *Proc. Natl. Acad. Sci. USA* 112 (34), 10798–10803. doi:10.1073/pnas.1503129112

Li, W., Zhu, Z., Chern, M., Yin, J., Yang, C., Ran, L., et al. (2017). A Natural Allele of a Transcription Factor in rice Confers Broad-Spectrum Blast Resistance. *Cell* 170 (1), 114–126. doi:10.1016/j.cell.2017.06.008

Liu, J., Yu, Y., Li, F., Zhang, X., and Xiang, J. (2014a). A New ALF from *Litopenaeus vannamei* and its SNPs Related to WSSV Resistance. *Chin. J. Ocean. Limnol.* 32 (6), 1232–1247. doi:10.1007/s00343-015-4010-4

Liu, J., Yu, Y., Li, F., Zhang, X., and Xiang, J. (2014b). A New Anti-lipopolysaccharide Factor (ALF) Gene with its SNP Polymorphisms Related to WSSV-Resistance of *Litopenae vannamei*. *Fish Shellfish Immunol.* 39 (1), 24–33. doi:10.1016/j.fsi.2014.04.009

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta}$CT Method. *Methods* 25 (4), 402–408. doi:10.1006/meth.2001.1262

Mai, H. N., Caro, L. F. A., Cruz-Flores, R., White, B. N., and Dhar, A. K. (2021). Differentially Expressed Genes in Hepatopancreas of Acute Hepatopancreatic Necrosis Disease Tolerant and Susceptible Shrimp (*Penaeus vannamei*). *Front. Immunol.* 12, 634152. doi:10.3389/fimmu.2021.634152

Maralit, B. A., Jaree, P., Boonchuen, P., Tassanakajon, A., and Somboonwiwat, K. (2018). Differentially Expressed Genes in Hemocytes of *Litopenaeus vannamei* Challenged with *Vibrio parahaemolyticus* AHPND (VP_{AHPND}) and VP_{AHPND} Toxin. *Fish Shellfish Immunol.* 81, 284–296. doi:10.1016/j.fsi.2018.06.054

Marston, S. (2018). The Molecular Mechanisms of Mutations in Actin and Myosin that Cause Inherited Myopathy. *Ijms* 19 (7), 2020. doi:10.3390/ijms19072020

Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a Major QTL for Resistance to Infectious Pancreatic Necrosis in Atlantic salmon (*Salmo salar*): Population-Level Associations between Markers and Trait. *BMC Genomics* 10 (1), 368. doi:10.1186/1471-2164-10-368

Ng, T. H., Lu, C. W., Lin, S. S., Chang, C. C., Tran, L. H., Chang, W. C., et al. (2018). The Rho Signalling Pathway Mediates the Pathogenicity of AHPND-causing *V. parahaemolyticus* in Shrimp. *Cell Microbiol.* 20 (8), 12. doi:10.1111/cmi.12849

Qi, C., Wang, L., Liu, M., Jiang, K., Wang, M., Zhao, W., et al. (2017). Transcriptomic and Morphological Analyses of *Litopenaeus vannamei* Intestinal Barrier in Response to *Vibrio paraheamolyticus* Infection Reveals Immune Response Signatures and Structural Disruption. *Fish Shellfish Immunol.* 70, 437–450. doi:10.1016/j.fsi.2017.09.004

Qin, Z., Babu, V. S., Wan, Q., Zhou, M., Liang, R., Muhammad, A., et al. (2018). Transcriptome Analysis of Pacific White Shrimp (*Litopenaeus vannamei*) Challenged by *Vibrio parahaemolyticus* Reveals Unique Immune-Related Genes. *Fish Shellfish Immunol.* 77, 164–174. doi:10.1016/j.fsi.2018.03.030

Rao, R., Bing Zhu, Y., Alinejad, T., Tiruvayipati, S., Lin Thong, K., Wang, J., et al. (2015). RNA-seq Analysis of *Macrobrachium rosenbergii* Hepatopancreas in Response to *Vibrio parahaemolyticus* Infection. *Gut Pathog.* 7 (1), 6. doi:10.1186/s13099-015-0052-6

Ren, X., Zhang, Y., Liu, P., and Li, J. (2019). Comparative Proteomic Investigation of *Marsupenaeus japonicus* Hepatopancreas Challenged with *Vibrio parahaemolyticus* and White Spot Syndrome Virus. *Fish Shellfish Immunol.* 93, 851–862. doi:10.1016/j.fsi.2019.08.039

Sandberg, R., Yasuda, R., Pankratz, D. G., Carter, T. A., Del Rio, J. A., Wodicka, L., et al. (2000). Regional and Strain-specific Gene Expression Mapping in the Adult Mouse Brain. *Proc. Natl. Acad. Sci.* 97 (20), 11038–11043. doi:10.1073/pnas.97.20.11038

Santos, C. A., Andrade, S. C. S., Teixeira, A. K., Farias, F., Guerrelhas, A. C., Rocha, J. L., et al. (2021). Transcriptome Differential Expression Analysis Reveals the Activated Genes in *Litopenaeus vannamei* Shrimp Families of superior Growth Performance. *Aquaculture* 531, 735871. doi:10.1016/j.aquaculture.2020.735871

Santos, C. A., Andrade, S. C. S., Teixeira, A. K., Farias, F., Kurkjian, K., Guerrelhas, A. C., et al. (2018). *Litopenaeus vannamei* Transcriptome Profile of Populations Evaluated for Growth Performance and Exposed to White Spot Syndrome Virus (WSSV). *Front. Genet.* 9, 120. doi:10.3389/fgene.2018.00120

Sarathi, M., Ahmed, V. P. I., Venkatesan, C., Balasubramanian, G., Prabavathy, J., and Hameed, A. S. S. (2007). Comparative Study on Immune Response of *Fenneropenaeus indicus* to *Vibrio alginolyticus* and White Spot Syndrome Virus. *Aquaculture* 271, 8–20. doi:10.1016/j.aquaculture.2007.07.002

Shi, X., Meng, X., Kong, J., Luan, S., Luo, K., Cao, B., et al. (2018). Transcriptome Analysis of 'Huanghai No. 2' *Fenneropenaeus chinensis* Response to WSSV Using RNA-Seq. *Fish Shellfish Immunol.* 75, 132–138. doi:10.1016/j.fsi.2018.01.045

Sitbon, Y. H., Yadav, S., Kazmierczak, K., and Szczesna-Cordary, D. (2019). Insights into Myosin Regulatory and Essential Light Chains: a Focus on Their Roles in Cardiac and Skeletal Muscle Function, Development and Disease. *J. Muscle Res. Cel Motil* 41, 313–327. doi:10.1007/s10974-019-09517-x

Soonthornchai, W., Chaiyapechara, S., Klinbunga, S., Thongda, W., Tangphatsornruang, S., Yoocha, T., et al. (2016). Differentially Expressed Transcripts in Stomach of *Penaeus monodon* in Response to AHPND Infection. *Dev. Comp. Immunol.* 65, 53–63. doi:10.1016/j.dci.2016.06.013

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene Set Enrichment Analysis: a Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. *Proc. Natl. Acad. Sci.* 102 (43), 15545–15550. doi:10.1073/pnas.0506580102

Sun, J.-J., Lan, J.-F., Shi, X.-Z., Yang, M.-C., Niu, G.-J., Ding, D., et al. (2016). β-Arrestins Negatively Regulate the Toll Pathway in Shrimp by Preventing Dorsal Translocation and Inhibiting Dorsal Transcriptional Activity. *J. Biol. Chem.* 291 (14), 7488–7504. doi:10.1074/jbc.M115.698134

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks. *Nat. Protoc.* 7 (3), 562–578. doi:10.1038/nprot.2012.016

Velázquez-Lizárraga, A. E., Juárez-Morales, J. L., Racotta, I. S., Villarreal-Colmenares, H., Valdes-Lopez, O., Luna-González, A., et al. (2019). Transcriptomic Analysis of Pacific White Shrimp (*Litopenaeus vannamei*, Boone 1931) in Response to Acute Hepatopancreatic Necrosis Disease Caused by *Vibrio parahaemolyticus*. *PLoS ONE* 14 (8), e0220993. doi:10.1371/journal.pone.0220993

Wang, L., Fan, C., Liu, Y., Zhang, Y., Liu, S., Sun, D., et al. (2014). A Genome Scan for Quantitative Trait Loci Associated with *Vibrio anguillarum* Infection Resistance in Japanese Flounder (*Paralichthys olivaceus*) by Bulked Segregant Analysis. *Mar. Biotechnol.* 16 (5), 513–521. doi:10.1007/s10126-014-9569-9

Wang, P.-H., Wan, D.-H., Gu, Z.-H., Deng, X.-X., Weng, S.-P., Yu, X.-Q., et al. (2011). *Litopenaeus vannamei* Tumor Necrosis Factor Receptor-Associated Factor 6 (TRAF6) Responds to *Vibrio alginolyticus* and White Spot Syndrome Virus (WSSV) Infection and Activates Antimicrobial Peptide Genes. *Dev. Comp. Immunol.* 35 (1), 105–114. doi:10.1016/j.dci.2010.08.013

Wang, Q., Yu, Y., Zhang, Q., Zhang, X., Huang, H., Xiang, J., et al. (2019). Evaluation on the Genomic Selection in *Litopenaeus vannamei* for the Resistance against *Vibrio parahaemolyticus*. *Aquaculture* 505, 212–216. doi:10.1016/j.aquaculture.2019.02.055

Woldemariam, N. T., Agafonov, O., Sindre, H., Høyheim, B., Houston, R. D., Robledo, D., et al. (2020). miRNAs Predicted to Regulate Host Anti-viral Gene Pathways in IPNV-Challenged Atlantic salmon Fry Are Affected by Viral Load, and Associated with the Major IPN Resistance QTL Genotypes in Late Infection. *Front. Immunol.* 11, 2113. doi:10.3389/fimmu.2020.02113

Yu, Y., Liu, J., Li, F., Zhang, X., Zhang, C., and Xiang, J. (2017). Gene Set Based Association Analyses for the WSSV Resistance of Pacific White Shrimp *Litopenaeus vannamei*. *Sci. Rep.* 7, 40549. doi:10.1038/srep40549

Zhang, Q., Yu, Y., Wang, Q., Liu, F., Luo, Z., Zhang, C., et al. (2019). Identification of Single Nucleotide Polymorphisms Related to the Resistance against Acute Hepatopancreatic Necrosis Disease in the Pacific White Shrimp *Litopenaeus*

vannamei by Target Sequencing Approach. *Front. Genet.* 10, 11. doi:10.3389/fgene.2019.00700

Zhang, X., Yuan, J., Sun, Y., Li, S., Gao, Y., Yu, Y., et al. (2019). Penaeid Shrimp Genome Provides Insights into Benthic Adaptation and Frequent Molting. *Nat. Commun.* 10 (1), 356. doi:10.1038/s41467-018-08197-4

Zheng, Z., Wang, F., Aweya, J. J., Li, R., Yao, D., Zhong, M., et al. (2018). Comparative Transcriptomic Analysis of Shrimp Hemocytes in Response to Acute Hepatopancreas Necrosis Disease (AHPND) Causing *Vibrio parahemolyticus* Infection. *Fish Shellfish Immunol.* 74, 10–18. doi:10.1016/j.fsi.2017.12.032

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 in Genetics

ORIGINAL RESEARCH
published: 03 January 2022
doi: 10.3389/fgene.2021.792666

Construction of Genetic Linkage Maps From a Hybrid Family of Large Yellow Croaker (*Larimichthys crocea*)

Xinxiu Yu[1,2], Rajesh Joshi[3], Hans Magnus Gjøen[1], Zhenming Lv[2] and Matthew Kent[1]*

[1]*Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, As, Norway,* [2]*National Engineering Research Centre of Marine Facilities Aquaculture, Marine Science and Technology College, Zhejiang Ocean University, Zhoushan, China,* [3]*GenoMar Genetics AS, Oslo, Norway*

Consensus and sex-specific genetic linkage maps for large yellow croaker (*Larimichthys crocea*) were constructed using samples from an F_1 family produced by crossing a Daiqu female and a Mindong male. A total of 20,147 single nucleotide polymorphisms (SNPs) by restriction site associated DNA sequencing were assigned to 24 linkage groups (LGs). The total length of the consensus map was 1757.4 centimorgan (cM) with an average marker interval of 0.09 cM. The total length of female and male linkage map was 1533.1 cM and 1279.2 cM, respectively. The average female-to-male map length ratio was 1.2 ± 0.23. Collapsed markers in the genetic maps were re-ordered according to their relative positions in the *ASM435267v1* genome assembly to produce integrated genetic linkage maps with 9885 SNPs distributed across the 24 LGs. The recombination pattern of most LGs showed sigmoidal patterns of recombination, with higher recombination in the middle and suppressed recombination at both ends, which corresponds with the presence of sub-telocentric and acrocentric chromosomes in the species. The average recombination rate in the integrated female and male maps was respectively 3.55 cM/Mb and 3.05 cM/Mb. In most LGs, higher recombination rates were found in the integrated female map, compared to the male map, except in LG12, LG16, LG21, LG22, and LG24. Recombination rate profiles within each LG differed between the male and the female, with distinct regions indicating potential recombination hotspots. Separate quantitative trait loci (QTL) and association analyses for growth related traits in 6 months fish were performed, however, no significant QTL was detected. The study indicates that there may be genetic differences between the two strains, which may have implications for the application of DNA-information in the further breeding schemes.

Keywords: large yellow croaker, RAD sequencing, linkage map, collinearity, recombination rate

OPEN ACCESS

Edited by:
Alexandre Wagner Silva Hilsdorf, University of Mogi das Cruzes, Brazil

Reviewed by:
Changxu Tian, Guangdong Ocean University, China
Zexi Cai, Aarhus University, Denmark

****Correspondence:***
Xinxiu Yu
xinxiu.yu@nmbu.no

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

Received: *11 October 2021*
Accepted: *17 November 2021*
Published: *03 January 2022*

Citation:
*Yu X, Joshi R, Gjøen HM, Lv Z and Kent M (2022) Construction of Genetic Linkage Maps From a Hybrid Family of Large Yellow Croaker (*Larimichthys crocea*). Front. Genet. 12:792666. doi: 10.3389/fgene.2021.792666*

INTRODUCTION

Large yellow croaker (*Larimichthys crocea*) has become an important aquaculture species in southeast China, where Mindong and Daiqu are the two major strains farmed. Artificial breeding of the Mindong strain started in 1985, while the Daiqu strain has been bred since 1999 (Chen et al., 2018). In 2019, the total production of large yellow croaker exceeded 220,000 tons and accounted for more than 12% of the cultured marine fish production of China (Yu X. J. et al., 2020).

The production is supported by several breeding programs, of which the majority are based on classical basic selection methods, like phenotypic selection (Chen et al., 2018). However, modern breeding approaches, such as marker assisted selection (MAS), can further enhance the genetic gain for economically important traits. MAS can greatly increase the efficiency if a sufficiently large QTL is detected, typically through QTL linkage mapping and association studies (Zenger et al., 2019). An outstanding example of MAS applied in aquaculture was the discovery of a QTL imparting resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon, accounting for about 80% of the total variation in this trait (Moen et al., 2009; Houston et al., 2012). Information from the QTL was used in selective breeding to generate IPN resistant fish, which now dominate production in Norway, leading to a remarkable reduction in IPN outbreaks (Norris, 2017). Subsequent studies have provided functional genomics data indicating that mutations in the epithelial cadherin gene (cdh1) affect virion internalization (Moen et al., 2015), demonstrating the power of genomic tools to help reveal the mechanistic basis for important traits. Although MAS can be useful for some traits where major QTLs have been identified, most traits of economic importance in aquaculture species (i.e., production traits) are assumed to be polygenic, and often have low-to-moderate heritabilities (Zenger et al., 2019). As a result, application of MAS to improve these complex traits may be inefficient. For such polygenic traits, *genomic selection* is a viable alternative, based on genomic breeding values predicted on a genome-wide scale, allowing even small QTLs to contribute (Meuwissen et al., 2001). For large yellow croaker, the estimates of heritability for body weight (0.31 ± 0.06), body length (0.33 ± 0.06) and body height (0.41 ± 0.07) in 6 months fish, and the genetic correlations between them ranged from 0.74 to 0.95 (Yu X. X. et al., 2020).

High-throughput sequencing has transformed genetics by making it relatively easy to generate genome-wide genetic marker datasets, which are a prerequisite for QTL identification in MAS. Significant progress was made through the discovery of cost-effective restriction-site associated DNA sequencing (RADseq) based strategies (Baird et al., 2008). RADseq can generate medium density SNP resources and has been successfully used in various fish species for genetic linkage maps, QTL analysis and population genetics (Davey and Blaxter, 2010), e.g., in Atlantic salmon (Houston et al., 2012; Gonen et al., 2014), channel catfish (Li et al., 2014) and Nile tilapia (Palaiokostas et al., 2013).

Several genetic linkage maps for large yellow croaker have been developed using different approaches (**Supplementary Table S1**). The first two genetic linkage maps made publicly available were constructed using amplified fragment length polymorphism (AFLP; Ning et al., 2007) and simple sequence repeats (SSR; Ye et al., 2014). However, next-generation sequencing technologies have made detection of large numbers of genome-wide SNP markers relatively easy, and Ao et al. (2015) constructed a SNP genetic linkage map with a total length of 5451.3 cM using RADseq, while Xiao et al. (2015) constructed a genetic map of 2,632 cM using RNA sequencing (RNAseq) of expressed genes. More recently, Kong et al. (2019) constructed a double-digest restriction-site associated DNA (ddRAD) based genetic map using 5261 SNPs with a total length of 1885.67 cM. Despite using different approaches, these SNP linkage maps have one thing in common, as they were all developed using only the Mindong strain.

Daiqu strain of large yellow croaker has been successfully cultured since 1999 and the aquaculture production is on an industrial scale (Chen et al., 2018). Most consumers prefer lean large yellow croaker, and the body shape has become an important economic trait (Dong et al., 2019). The Daiqu strain has better performance for this trait, as the ratio of body length and body height is significantly higher than for the Mindong strain (Huang et al., 2006). The Daiqu strain also has later sexual maturation and better tolerance to lower temperatures than the Mindong strain (Liu and Mitcheson, 2008; Miao et al., 2014). The offspring from a crossing between Mindong and Daiqu displayed significant heterosis in body shape and growth of fish after 526 days (Li et al., 2010). Our study therefore sought to develop a genetic linkage map in a crossed (F_1) family arising from these strains.

The aim of this study was to construct consensus and sex-specific linkage maps based on a hybrid family from the Daiqu and Mindong strains using RADseq, to compare the linkage map to the latest physical map (*ASM435267v1*), and to perform a QTL analysis and association analysis for growth-related traits.

MATERIALS AND METHODS

Mapping Family

A female (F_3 of wild Daiqu strain, from an aquaculture farm in Xiangshan, Zhejiang province) and a male (approx. F_{10} of wild Mindong strain, from an aquaculture farm in Fuding, Fujian province) large yellow croaker were crossed to generate a fullsib family (Yu X. et al., 2017). One-hundred and twenty offspring were randomly selected at 6 months, and the following growth traits were recorded: body weight (BW), body length (BL) and body height (BH). Fin clips were preserved in 99% ethanol and sent to BGI Genomics Company (Shenzhen, China) for sequencing.

RAD Sequencing and SNP Calling

Library preparation was performed by BGI according to Baird et al. (2008). In brief, individual genomic DNA samples were digested using the restriction enzyme *Pst I*, and the resulting fragments were ligated to a double-stranded Illumina sequencing primer containing a sample-specific barcode sequence. Libraries were then pooled and sheared by sonication, and fragments from 300–500 bp were separated by agarose gel electrophoresis and purified before ligating a Y-adapter to the sheared ends. Fragments including both barcode and Y-adapters were amplified with PCR to generate the final RAD libraries, which were then sequenced using a *Hiseq2000* platform to produce paired-end reads.

Raw reads were processed by BGI using the *Reseqtools* software package (https://github.com/BGI-shenzhen/Reseqtools) to remove adapter sequences and low-quality

reads, and to de-multiplex the pool. The retained reads were analysed and genotyped using *Stacks* (Catchen et al., 2013) and in-house analysis pipelines, and RAD-tags with too low (<2) or too high (>100) sequencing coverage were excluded.

SNP Filtering and Linkage Map Construction

SNPs missing in >10% of samples and minor allele frequency (MAF) < 0.05 were excluded using the PLINK software (Purcell et al., 2007). Markers were individually tested against the expected segregation ratio, based on parental genotypes, and those showing significant segregation distortion ($p < 0.05$, χ^2 test) were removed by PLINK.

The remaining SNPs were used to generate consensus and sex-specific maps using Lep-MAP2 software (Rastas et al., 2013). All SNP markers that passed filtering (n = 20,186) were used to produce the consensus map, while those markers polymorphic in the father (n = 11,684) or mother (n = 11,838) were used to construct their respective sex-specific linkage maps. LGs were developed using the *separate chromosomes* module, with a logarithm of odds (LOD) score ranging from 1 to 20. A LOD score of 9, which gave 24 LGs and the lowest number of single markers, was finally selected. The option *sizeLimit =100* was used to generate linkage groups of size ≥100 markers. The module *JoinSingles* could not assign any of the singular markers to any of the 24 LGs. Eventually, 20,147 SNPs were ordered using the *OrderMarkers* module, which assign the markers with paternal or maternal positions for the sex-specific maps. The option *sexAverage = 1* was applied during execution of *OrderMarkers* to get positions for the consensus map. To avoid the map distances being too long, especially when the number of markers per chromosome was much higher than the number of individuals, the parameter *minError = 0.15* was used. Finally, the *Kosambi* mapping function was used to calculate genetic distance between markers. The LG were numbered by the SNP size of each LG (i.e., the LG with the largest SNP number was labelled LG1). Illustrations of the consensus and sex-specific linkage maps were drawn using MapChart 2.32 (Voorrips, 2002).

QTL Analysis and Association Analysis

QTL analysis was initially performed using the QTL *IciMapping* software by the option *inclusive composite interval mapping with an additive effect* (ICIM-ADD) (Li et al., 2015; Meng et al., 2015). The LOD threshold for QTL significance of each trait was determined by a permutation test (1,000 replications) with a genome-wide significance level of 0.05. The permutation threshold method for QTL mapping estimates the null distribution of the genome-wide maximum LOD score by shuffling the phenotypes relative to the genotype data, breaking the association between the phenotype and the genotypes (Churchill and Doerge, 1994). The genome-wide LOD thresholds are calculated based on the 1-α quantiles of the genome-wide maximum LOD scores obtained from the permutations, where α is the significance level ($\alpha = 0.05$ in our case).

As a complementary method for QTL mapping, a genome-wide association study (GWAS) was performed using SNPs subjected to a more stringent quality filtering than that was applied for linkage mapping to ensure a high QTL identification accuracy. Using PLINK, individuals displaying more than 5% missing genotypes were removed. Also, SNPs were removed in cases where missing genotypes >5% across samples and Hardy-Weinberg p value (Fishers exact test) < 10^{-9}. The final SNP set used for GWAS thus included 16,570 SNPs from 74 individuals.

The genome-wide association analysis was performed using a mixed linear model equation on BW, BL and BH by the *Genome-wide Complex Trait Analysis* (GCTA) program, with the *-mlma* function (Yang et al., 2011). The following model was used:

$$\mathbf{y} = \mathbf{a} + \mathbf{bx} + \mathbf{g} + \mathbf{e}$$

where **y** is the phenotypes (BW, BL, BH), a is the overall mean for each trait, **b** is the additive genetic effect of the candidate SNP to be tested for association, x is the incidence matrix for the candidate SNPs, **g** is the polygenic effect and **e** is the vector of random residual effects.

SNPs were considered genome wide significant when exceeding the Bonferroni threshold for multiple testing ($\alpha = 0.05$) of 0.05/tg = 3.017502×10^{-6}, where tg = 16,570 (total number of genome-wide SNPs); and SNPs were graded as chromosome-wide significant when Bonferroni threshold for multiple testing ($\alpha = 0.05$) surpassed 0.05/tc = 7.246377×10^{-5}, where tc = 690 (average number of SNPs per chromosome). The genome-wide significant threshold used in this study was $p \leq 3.017502 \times 10^{-6}$ (–log10 (P) = 5.52), while chromosome-wide significant threshold was $p \leq 7.246377 \times 10^{-5}$ (–log10 (P) = 4.14). SNPs were visualised along the linkage groups using the Manhattan function in the R package QQMAN (Turner, 2014).

Collinearity Analysis: Genetic vs. Physical Map

To explore the level of agreement between our consensus genetic map and a recently published physical map, the large yellow croaker assembly, *ASM435267v1* (GenBank ID GCA_004352675.1), the RAD-tag sequences (82 bp) from the consensus linkage map were aligned to *ASM435267v1* using BLASTN (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with the following parameters: *expect value* $e \leq 1 \times 10^{-15}$, *identity ≥ 95%, matched length ≥ 81 bp, mismatches ≤ 1* and *gap open = 0*. If a query sequence hit two or more loci in the physical assembly and the difference between the 1st and 2nd smallest e-values was greater than 10^3, the 1st smallest e-value was chosen to define the hit. Finally, 9885 SNPs from the consensus map hit the physical map.

The relative positioning of RAD sequences in the genetic map and the physical map were graphically presented using *shinyCircos* (Yu Y. et al., 2017). The marker positions on genetic map were multiplied by 4×10^5 for better visualisation of the Circos plot.

Adjusting Genetic Maps Based on the Physical Map

The collinearity analysis highlighted 107 SNPs whose assignment to LGs disagreed with their physical assignment to chromosomes.

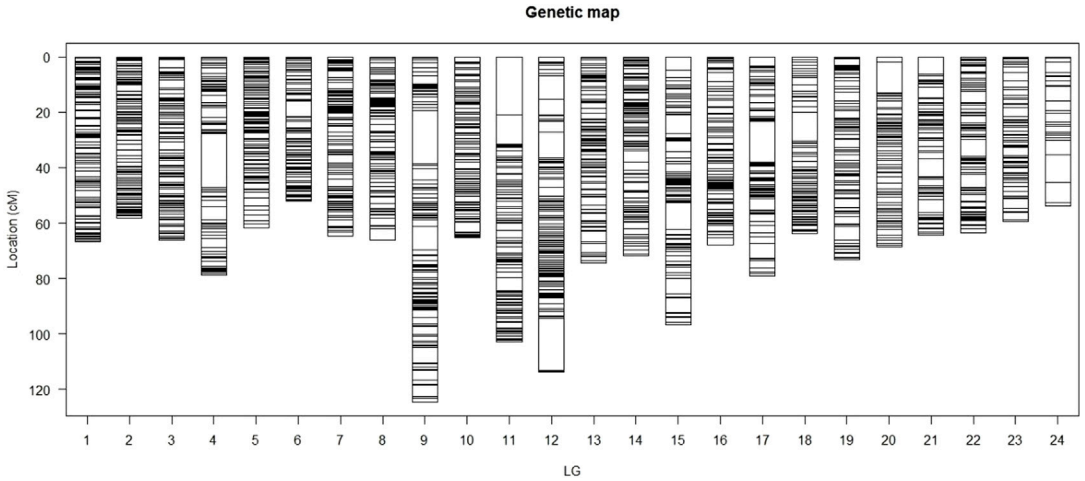

FIGURE 1 | | The consensus linkage map for large yellow croaker. The dark bands show the density of the SNPs in the region of the LGs, whereas white bands show the regions with no SNPs.

Examples of this were seen in all LGs and, when detected, SNPs were reordered according to the physical map. The adjusted SNP order in each linkage group was used as an input of *evaluateOrder* option in the *OrderMarkers* module of Lep-MAP2, and the genetic distances were recalculated using the *Kosambi* mapping function. The integrated consensus and sex-specific linkage maps were drawn using *MapChart 2.32* (Voorrips, 2002).

Scatter plots were generated between the integrated linkage maps in cM distances and the physical map in Mb distances by using the *ggplot2* package in R. Recombination rates throughout the genome in the integrated female and male genetic maps were estimated using *MareyMap* online (Siberchicot et al., 2017) with a computed sliding window size of 3.37 Mb. The threshold markers number in a window was set to 8, the default value. The recombination rate changes throughout the genome in the integrated female and male maps were visualised by using the *ggplot2* package in R.

RESULTS

Sequencing and SNP Filtering

Approximately 1.9 billion reads were produced after sequencing two parents and 120 offspring, with each individual contributing roughly 15 ± 2.9 million reads. After reads filtering and RAD-tag SNP detection, approximately 370,000 variants were detected within each individual. The average heterozygosity rate was 32.5%. After filtering for segregation errors, MAF and missing genotypes, a final set of 20,186 SNP markers was used for linkage map construction.

Linkage Map Construction

The SNPs were assigned to 24 LGs, in accordance with the haploid chromosome number (Lou et al., 2015). The consensus map (**Figure 1**; **Table1**) covered 1757.4 cM, with individual linkage group lengths ranging from 51.9 cM (LG6) to 124.6 cM (LG9). The number of markers per linkage group varied from 243 to 1,230, with an average genetic distance between markers of 0.09 cM and a standard deviation of 0.037 (**Table1**).

Separate male and female maps were constructed using segregating (heterozygous) markers from each parent. The total length of the male linkage map was 1,279 cM, and the total length of the female map was 1,533 cM (**Table1**, **Supplementary Figure S1**). In the female map, LG length ranged from 39.8 cM (LG22) to 86.7 cM (LG3), and the SNP number per LG varied from 114 to 783. In the male map, the length of each LG varied from 43.1 cM (LG2) to 83.5 cM (LG4), and the SNP number in each LG varied from 158 to 745. The average distance between markers for female and male is thus 0.11 and 0.13 cM, respectively. The female-to-male length ratio ranged from 0.7 (LG22) to 1.6 (LG3), with an average of 1.2 ± 0.23; most LGs in the female map were larger than those in the male map, with the exceptions of LG4, LG19, LG21 and LG24.

QTL Analysis and Association Analysis for Growth Traits

The growth traits, BW, BL and BH, recorded in 120 offspring at 6 months of age, are presented in **Table 2**. In the QTL analysis, the LOD threshold used was 8.81 for BW, 7.35 for BL, and 18.71 for BH. However, no QTL was above the LOD threshold for any of the growth traits, BW, BL or BH (**Supplementary Figure S3**). In the GWAS analysis, the estimated genomic heritabilities for the three traits were close to zero (**Table 3**). A total of 16,570 SNPs from 74 recorded individuals were used, however, no SNPs crossed the genome or chromosome-wide significant level (**Supplementary Figure S4**).

Collinearity Analysis

In total, 9885 SNPs from the consensus map hit the physical map (*ASM435267v1*), but there were 107 SNPs hitting non-

TABLE 1 | | Key figures for the genetic linkage maps of large yellow croaker.

LG	Consensus			Female		Male	
	No. of markers	Size (cM)	Average distance (cM)	No. of markers	Size (cM)	No. of markers	Size (cM)
1	1,230	66.8	0.05	613	78.2	745	51.8
2	1,125	58.3	0.05	783	65.7	716	43.1
3	1,029	66.0	0.06	660	86.7	574	55.5
4	1,000	78.7	0.08	654	66.8	503	83.5
5	997	61.8	0.06	626	72.6	629	49.4
6	973	51.9	0.05	543	54.2	548	43.8
7	967	64.6	0.07	532	67.2	558	50.1
8	949	66.3	0.07	458	70.2	647	60.4
9	899	124.6	0.14	609	68.3	375	58.0
10	875	65.2	0.07	484	68.6	505	49.4
11	850	102.8	0.12	475	64.4	468	51.8
12	837	113.7	0.14	533	68.6	410	59.3
13	831	74.4	0.09	480	58.5	480	52.8
14	817	71.9	0.09	589	71.8	377	50.7
15	814	96.6	0.12	297	70.4	589	52.9
16	791	68.0	0.09	387	61.2	526	55.2
17	789	79.0	0.10	417	70.4	476	45.6
18	789	63.9	0.08	616	55.2	566	48.6
19	787	73.2	0.09	589	54.2	283	56.6
20	741	68.4	0.09	539	54.2	358	46.1
21	693	64.5	0.09	378	50.5	392	55.3
22	689	63.6	0.09	237	39.8	528	58.0
23	432	59.3	0.14	225	65.7	273	50.5
24	243	53.9	0.22	114	49.8	158	50.8
Total	20147	1757.4	0.09	11838	1533.1	11684	1,279.2

TABLE 2 | Mean ± SD, range and coefficient of variation (CV) of growth traits at 6 months.

	Body weight, g	Body length, cm	Body height, cm
Mean ± SD	45.2 ± 13.8	13.8 ± 1.6	3.6 ± 0.5
Range	20.8–89.3	9.5–17.7	2.8–7.6
CV (%)	31	11	15

corresponding chromosomes. A collinearity analysis, comparing the consensus linkage and physical maps, was performed (**Figure 2**). The average correlation coefficient between the genetic map and the physical map was 0.78 ± 0.16 (**Supplementary Table S2**). Each LG matches well with its corresponding chromosome of the physical map, with an average matching percentage of 98.92 ± 1.5%. There were 7 LGs that showed no mismatch between the genetic map and the physical map; LG7, LG9, LG10, LG17, LG18, LG20 and LG24 (**Supplementary Table S3**).

Integration of Physical and Genetic Maps

SNP position information based on the *ASM435267v1* genome assembly was used to produce the physically informed consensus, female and male linkage maps (**Supplementary Figure S2**). A summary of the integrated maps is shown in **Supplementary Table S3**. A comparison of map positions between the integrated genetic and physical maps for different LGs is shown in **Figure 3**, in which most LGs exhibited sigmoidal patterns of recombination, with greater recombination rates toward the middle and low recombination rates toward the ends of the chromosomes. Large gaps or jumps can be seen in some of the plots, viewed from the *x*-axis or from the *y*-axis (**Supplementary Table S4**). Viewed from the *x*-axis, representing the physical position, large gaps were observed on LG24 (2.12 Mb) and LG1 (1.9 Mb), whereas viewed from the *y*-axis, representing the genetic position, large gaps were observed e.g., in LG4 (36.36 cM) of the integrated male map and in LG21 (50.37 cM) of the integrated female map. The markedly large jump downward in LG23 of the integrated consensus map was due to the fragmented linkage group LG23.1 assigned to LG23 in this case. Recombination rates of the three integrated maps are shown in **Supplementary Table S3**, and the recombination rate variation comparison of integrated female and male maps is visualised in **Figure 4**. The average recombination rate in the female was 3.55 cM/Mb whereas it in the male was 3.05 cM/Mb. The pattern of the recombination rates was different between male and female in some LGs, as there was a higher recombination rate for the male than for the female in the beginning of some LGs (e.g., LG09), whereas in other LGs the pattern was just opposite (e.g., LG20).

DISCUSSION

Linkage Map Construction and Collinearity Analysis

The total genetic length of the consensus linkage map in our study was 1757.4 cM. The genetic map length (1885.67 cM) using the Mindong strain only, found by Kong et al. (2019), is slightly larger than that of our study. However, the linkage map length, also

TABLE 3 | Estimates of variance components and heritability with standard errors (in parenthesis) using the genomic relationship matrix in GWAS by GCTA.

Traits	σ_g^2	σ_e^2	σ_P^2	Genomic h^2
BW	0.00022 (48.66)	194.5570 (42.94)	194.5572 (35.96)	0.000001 (0.25)
BL	0.000003 (0.69)	2.925069 (0.70)	2.925072 (0.52)	0.000001 (0.24)
BH	0 (0.08)	0.377716 (0.09)	0.37716 (0.07)	0.000001 (0.22)

σ_g^2, Genetic variance; σ_P^2, Phenotypic variance; σ_e^2, Residual variance; h^2, Heritability.

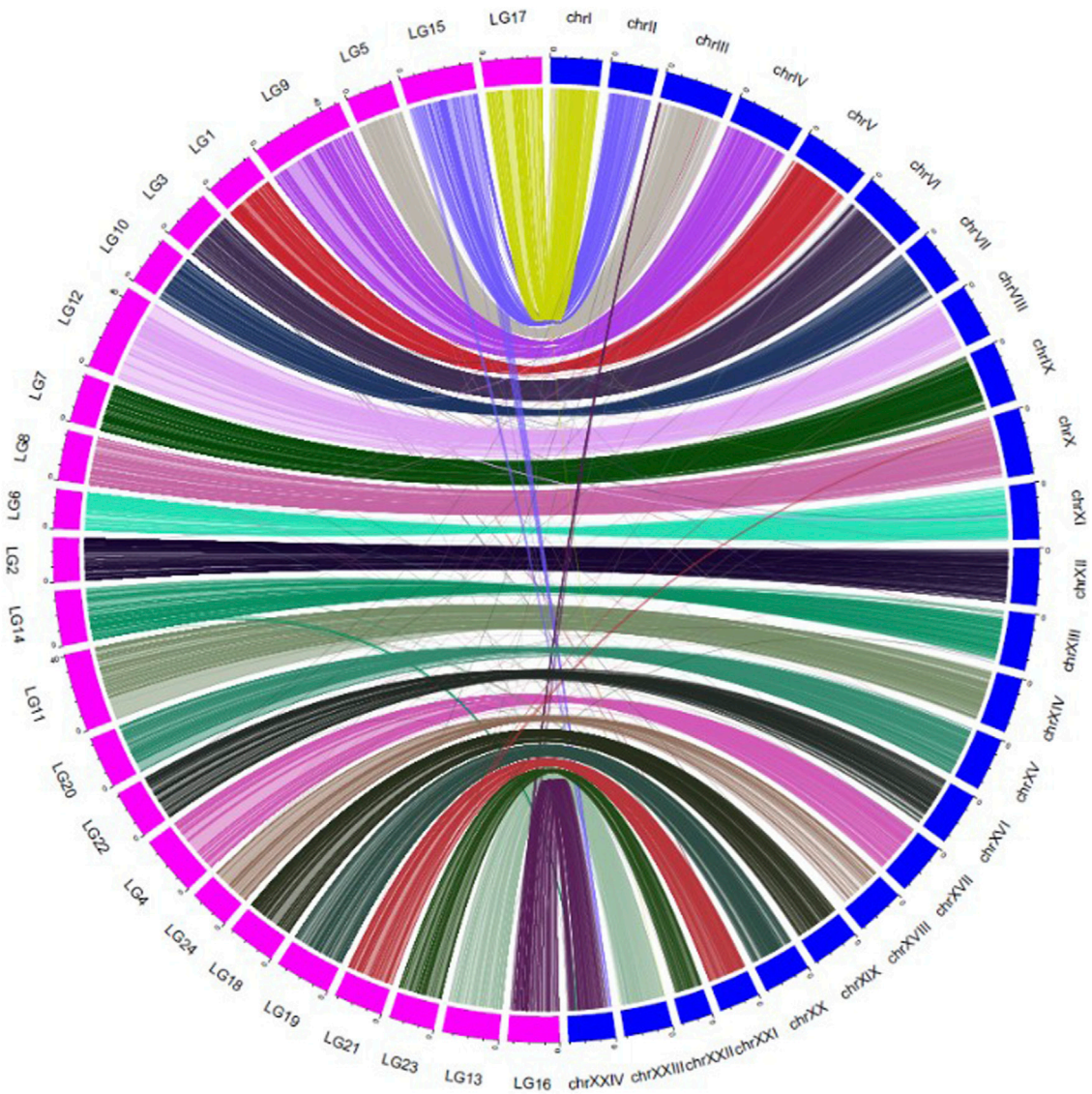

FIGURE 2 | Collinearity between the consensus genetic map (LG1-24) and the physical map (chr I-XXIV) of large yellow croaker.

using Mindong strain, found by Ao et al. (2015) was 5451.3 cM, is much larger than in our study. The differences in total genetic length could be caused by the mapping family used in our study, which was a cross between Mindong strain and Daiqu strain. Suppressed recombination rates have also been reported in rainbow-Yellowstone cutthroat trout (*Oncorhynchus clarkii bouvieri*) hybrids, there explained by chromosome rearrangements (Ostberg et al., 2013). And in a hybrid cross of Human Pathogenic Fungus, *Cryptococcus neoformans*, the linkage map length (197 cM) was much shorter than those (1,356.3 cM) observed in a single strain (Sun and Xu, 2007).

The average female-to-male map length ratio was 1.2 ± 0.23 in our study, indicating more recombination events happening in females, which is consistent with an earlier study in large yellow

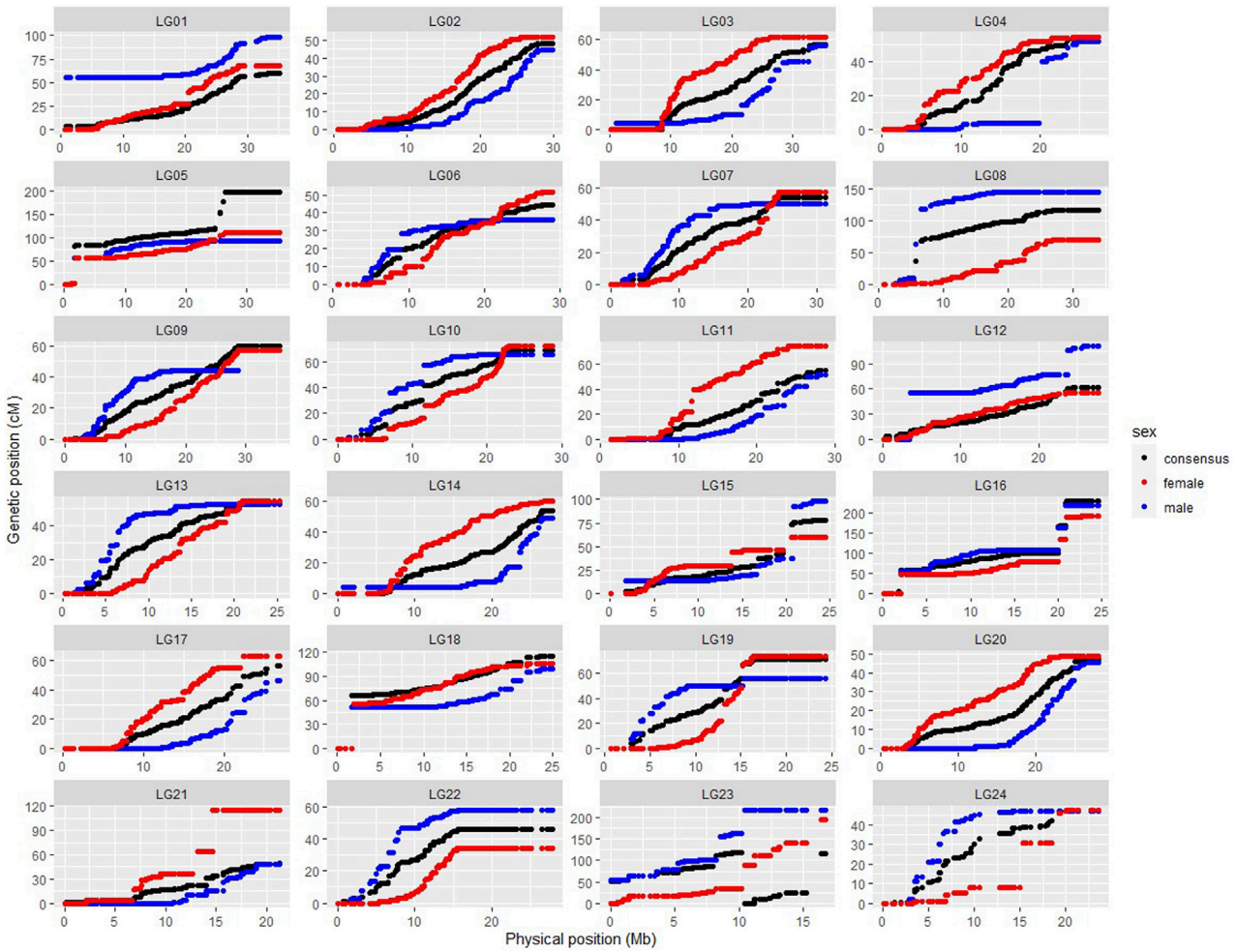

FIGURE 3 | Scatter plots showing SNP linkage map positions (cM; *y*-axis) versus physical positions (Mb; *x*-axis) for integrated female (red), male (blue) and consensus (black) genetic maps. (Fragmented linkage group LG23.1 was assigned to LG 23.)

croaker by Ning et al. (2007). The phenomenon of heterochiasmy, i.e., sex differences in recombination rates between the two sexes, has been found in many fish species. Higher recombination rate in female fish, as in our case, was also reported in Atlantic salmon (1.38) (Lien et al., 2011), gilthead sea bream (1.61) (Tsigenopoulos et al., 2014), Nile tilapia (1.2) (Joshi et al., 2018) and Gasterosteus sticklebacks (1.64) (Sardell et al., 2018), where in all cases it seems that the heterogametic sex has lower recombination rates. In our study, most LGs in the female map were larger than those in the male map, whereas the male map was larger in LG4, LG19, LG21 and LG24. Similar cases were also found in other fish, such as the gilthead seabream and Nile tilapia (Tsigenopoulos et al., 2014; Joshi et al., 2018). The molecular mechanisms for the sex differences in recombination rates are still not well understood. The differences may be caused by sexually antagonistic selection, meiotic drive in females, selection during the haploid phase of the life cycle, selection against aneuploidy, or mechanistic constraints; however, no single hypothesis can adequately explain the evolution of heterochiasmy in all cases (Sardell and Kirkpatrick, 2020).

Sigmodal patterns of recombination in large yellow croaker, with greater recombination rates toward the middle and lower recombination rates toward the ends, have also been seen in other species, like Nile tilapia (Joshi et al., 2018), whereas, salmon, channel catfish, etc (Li et al., 2014; Tsai et al., 2015) have shown opposite patterns, with higher recombination at the end of the LGs. The segments with little or no recombination may suggest possible location of centromeres. The karyotypes of large yellow croaker were earlier categorised into 10 pairs of sub-telocentric and 14 pairs of acrocentric chromosomes (Xu et al., 2017), implying that the centromeres are located at the end of the chromosomes, matching the low recombination rates seen towards the end of these LGs in our study. Recombination rate profiles within each LG also differed between males and female, with distinct regions containing potential recombination hotspots.

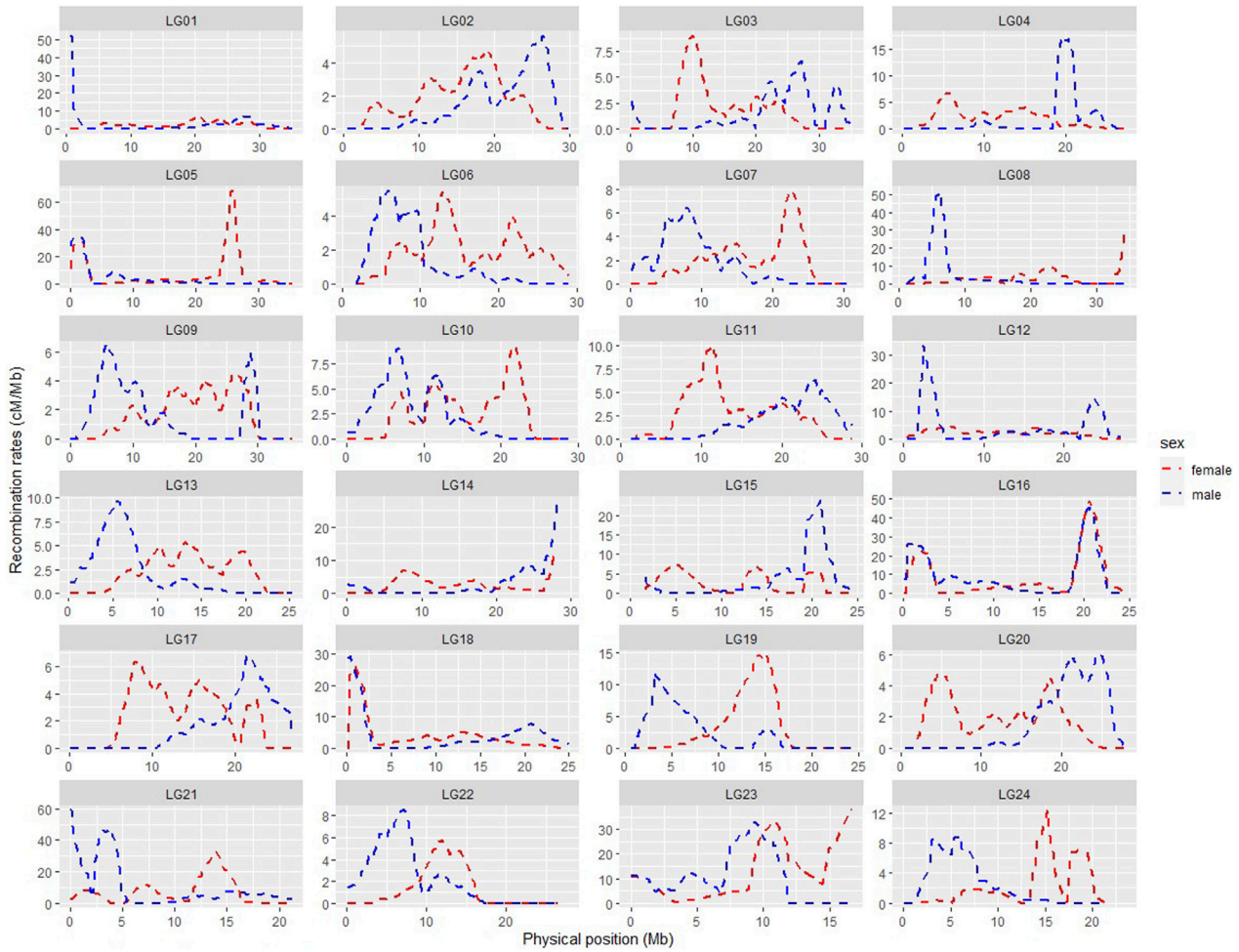

FIGURE 4 | Comparison of recombination rate variation throughout the genome in the integrated female (red) and male (blue) maps.

Physical gaps, as viewed from the *x*-axis in **Figure 3**, indicate lack of SNPs in these regions. One reason for these gaps could be massive repeat sites, unrecognisable by the *Pst I* enzyme during RADseq. Identifying additional markers with a different enzyme should thus help to fill these gaps. Another related reason could be the random and consequently partly uneven distribution of detected markers across the genomes, which is a disadvantage of RAD based technologies. Thus, RADseq usually generates medium density SNP linkage maps, leading to a low genome coverage (Robledo et al., 2017). Furthermore, the strain used in the *ASM435267v1* genome assembly, called DH2-L1, is a double haploid obtained by artificial gynogenesis from the Mindong strain only (Cai et al., 2010), whereas the population used in our genetic map is a cross between Daiqu strain and Mindong strain. The strain difference could thus be another reason for the physical gaps, as chromosomal rearrangements, including deletions, duplications, inversions, and translocations, could be different among strains.

Large jumps in some of the LGs were also viewed from the *y*-axis in **Figure 3**. These regions, with significantly elevated recombination rates, may be due to recombination hotspots, insufficient SNP coverage caused by the randomness of RAD sequencing explained above, and/or low level of polymorphism in the F_1 family. A similar problem of large intervals was also presented in the genetic linkage map of the small yellow croaker (*Larimichthys polyactis*), also from only one fullsib family (Liu et al., 2020). Thus, use of multiple fullsib or halfsib families should be preferred, as done for instance with the high-density linkage map developed in Nile tilapia using 41 fullsib families (Joshi et al., 2018).

QTL Analysis and GWAS for Growth Traits

QTL analysis and GWAS are two types of strategies to detect potential causal genes for quantitative traits. QTL analysis detect associations between marker intervals and phenotypes, while GWAS identifies associations between single DNA markers and phenotypes, and thus the two methods complement each other (Sonah et al., 2015).

For fish less than 10 months of age, the gonads are hard to assess only by naked eye observation and there is hardly any gender difference to use for sex determination (Wang and Cai, 2018). Thus, no gender information was available for the fish at

6 months in the present study. Growth differences have already been observed between the Mindong and Daiqu strains and some phenotype segregation may be expected in the F_1, but no significant QTLs or SNPs were detected in our QTL or GWAS analysis. This was probably due to the complex genetic nature of the three growth traits which generally have been found to be controlled by many genes, each with minor effects. Also, the power of QTL analysis and GWAS will often not be sufficient with only one test family, due to the categorical nature of QTLs, for which a significant variant may or may not be present in any given family. For instance, the highly significant QTL variant that induced high resistance to IPN virus in the study of Moen et al. (2009), was only present in ca 5 % of the breeding nucleus. The QTL plot of BH is close to the threshold by 1,000 permutations, while the Manhattan plot of BH is far from the suggestive threshold by Bonferroni correction, which has been reported to be overly conservative in some cases (Kaler and Purcell, 2019). Also, no SNPs were identified to be significantly associated with the BL/BD ratio or the BL/BH ratio in large yellow croaker (Dong et al., 2019; Zhou et al., 2019). This may be due to low power in all these studies, but the results correspond well with the assumed polygenic nature of these traits. However, Xiao et al. (2015) identified several potential QTLs for growth traits (total weight, total length and total height) by composite interval mapping using 72 individuals from one fullsib family. But LOD score significance thresholds were not given in the plots in this study.

Using one F_1 fullsib family, as in the present study, Kong et al. (2019) identified seven significant QTLs linked to white spot disease resistance. The probability of identifying the QTLs in disease resistance traits could be higher than in growth traits, as it is often found that they are controlled by some major QTLs (Fraslin et al., 2020). However, these studies, using one F_1 family, only provide preliminary results of QTL mapping, and studies involving a more representative sample of the breeding population are required to conduct a marker-assisted selection scheme. One fullsib family is thus not ideal for identifying candidate genes, and a larger sample size and more families should be used to improve the power and to reveal potential associations (Korte and Farlow, 2013).

CONCLUSION

A consensus genetic linkage map for large yellow croaker was constructed with 20,147 SNPs from RAD sequencing, based on an F_1 family from Mindong strain and Daiqu strain. The total length of the consensus map was 1757.4 cM with an average marker interval of 0.09 cM. The female-to-male linkage map length ratio was 1.2. The map was adjusted based on the physical map, and integrated consensus and sex-specific linkage maps were generated. The recombination pattern mostly showed sigmoidal pattern of recombination. In most LGs, higher recombination rates were found in the integrated female map, compared to the integrated male map. No significant QTLs for growth related traits in fish at 6 months were found, probably due to the low detection power in only one family and the polygenic and complex nature of growth traits that are controlled by many genes with minor efforts. The present study indicates that there may be genetic differences between the two strains Daiqu and Mindong, which may have implications for breeding programs using DNA-information in a future selection scheme.

DATA AVAILABILITY STATEMENT

The sequenced data used for SNP detection has been deposited in Sequence Read Archive in NCBI (PRJNA786283). 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://figshare.com/, doi: 10.6084/m9. figshare. 16779160.

ETHICS STATEMENT

The animal study was reviewed and approved by the Animal Care and Use Committee of Zhejiang Ocean University.

AUTHOR CONTRIBUTIONS

ZL, XY: Conceptualization. HG, MK, and RJ: Supervision. RJ, XY, and MK: Data curation and analysis. XY: Samples preparation, result visualization and writing the original draft. RJ, MK, and HG: Reviewing and Editing.

FUNDING

This study was financially supported by the Open Foundation from Marine Sciences in the Most Important Subjects of Zhejiang (No.20130204).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.792666/full#supplementary-material

REFERENCES

Ao, J., Li, J., You, X., Mu, Y., Ding, Y., Mao, K., et al. (2015). Construction of the High-Density Genetic Linkage Map and Chromosome Map of Large Yellow Croaker (*Larimichthys Crocea*). *Ijms* 16 (11), 26237–26248. doi:10.3390/ijms161125951

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 (10), e3376. doi:10.1371/journal.pone.0003376

Cai, M., Wu, Q., Liu, X., Yao, C., Chen, Q., and Wang, Z. (2010). Artificial Induction of Mito-Gynogenetic Diploids in Large Yellow Croaker (*Pseudosciaena Crocea*) by Hydrostatic Pressure. *Chin. J. Ocean. Limnol.* 28, 713–719. doi:10.1007/s00343-010-908510.1007/s00343-010-9085-3

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an Analysis Tool Set for Population Genomics. *Mol. Ecol.* 22 (11), 3124–3140. doi:10.1111/mec.12354

Chen, S., Su, Y., and Hong, W. (2018). "Aquaculture of the Large Yellow Croaker," in *Aquaculture in China* (Hoboken, New Jersey: Wiley), 297–308. doi:10.1002/9781119120759.ch3_10

Churchill, G. A., and Doerge, R. W. (1994). Empirical Threshold Values for Quantitative Trait Mapping. *Genetics* 138, 963–971. doi:10.1093/genetics/138.3.963

Davey, J. W., and Blaxter, M. L. (2010). RADSeq: Next-Generation Population Genetics. *Brief. Funct. Genomics* 9, 416–423. doi:10.1093/bfgp/elq031

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*). *Aquacult. Fish.* 4, 3–8. doi:10.1016/j.aaf.2018.05.001

Fraslin, C., Quillet, E., Rochat, T., Dechamp, N., Bernardet, J.-F., Collet, B., et al. (2020). Combining Multiple Approaches and Models to Dissect the Genetic Architecture of Resistance to Infections in Fish. *Front. Genet.* 11, 677. doi:10.3389/fgene.2020.00677

Gonen, S., Lowe, N. R., Cezard, T., Gharbi, K., Bishop, S. C., and Houston, R. D. (2014). Linkage Maps of the Atlantic salmon (*Salmo salar*) Genome Derived from RAD Sequencing. *BMC Genomics* 15, 166. doi:10.1186/1471-2164-15-166

Houston, R. D., Davey, J. W., Bishop, S. C., Lowe, N. R., Mota-Velasco, J. C., Hamilton, A., et al. (2012). Characterisation of QTL-Linked and Genome-wide Restriction Site-Associated DNA (RAD) Markers in Farmed Atlantic salmon. *BMC Genomics* 13, 244. doi:10.1186/1471-2164-13-244

Huang, L. M., Xie, Y. J., and Su, Y. Q. (2006). Studies on Genetic Diversities of Daiqu Stock and Min-Yue Stock. *Pseudosciaena Crocea. J. Xiamen Univ. (Natural Science)* 45 (6), 836–840.

Joshi, R., Árnyasi, M., Lien, S., Gjøen, H. M., Alvarez, A. T., and Kent, M. (2018). Development and Validation of 58K SNP-Array and High-Density Linkage Map in Nile tilapia (*O. niloticus*). *Front. Genet.* 9, 472. doi:10.3389/fgene.2018.00472

Kaler, A. S., and Purcell, L. C. (2019). Estimation of a Significance Threshold for Genome-wide Association Studies. *BMC Genomics* 20, 618. doi:10.1186/s12864-019-5992-7

Kong, S., Ke, Q., Chen, L., Zhou, Z., Pu, F., Zhao, J., et al. (2019). Constructing a High-Density Genetic Linkage Map for Large Yellow Croaker (*Larimichthys Crocea*) and Mapping Resistance Trait against Ciliate Parasite *Cryptocaryon Irritans*. *Mar. Biotechnol.* 21, 262–275. doi:10.1007/s10126-019-09878-x

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

Li, M.-y., Hu, Y.-z., Miao, L., Chen, J., Shi, Y.-h., Xue, L.-y., et al. (2010). Studies on the Growth Characteristics and Heterosis of Genealogies of *Pseudosciaena Crocea*. *J. Fish. China* 34 (6), 679–684. doi:10.3724/sp.j.1231.2010.06750

Li, S., Wang, J., and Zhang, L. (2015). Inclusive Composite Interval Mapping of QTL by Environment Interactions in Biparental Populations. *PLoS One* 10 (7), e0132414. doi:10.1371/journal.pone.0132414

Li, Y., Liu, S., Qin, Z., Waldbieser, G., Wang, R., Sun, L., et al. (2014). Construction of a High-Density, High-Resolution Genetic Map and its Integration with BAC-Based Physical Map in Channel Catfish. *DNA Res.* 22, 39–52. doi:10.1093/dnares/dsu038

Lien, S., Gidskehaug, L., Moen, T., Hayes, B. J., Berg, P. R., Davidson, W. S., et al. (2011). A Dense SNP-Based Linkage Map for Atlantic salmon (*Salmo salar*) Reveals Extended Chromosome Homeologies and Striking Differences in Sex-specific Recombination Patterns. *BMC Genomics* 12, 615. doi:10.1186/1471-2164-12-615

Liu, F., Zhan, W., Xie, Q., Chen, H., Lou, B., and Xu, W. (2020). A First Genetic Linage Map Construction and QTL Mapping for Growth Traits in *Larimichthys Polyactis*. *Sci. Rep.* 10, 11621. doi:10.1038/s41598-020-68592-0

Liu, M., and de Mitcheson, Y. S. (2008). Profile of a Fishery Collapse: Why Mariculture Failed to Save the Large Yellow Croakerfile of a Fishery Collapse Why Mariculture Failed to Save the Large Yellow Croaker. *Fish and Fihseries* 9, 219–242. doi:10.1111/j.1467-2979.2008.00278.x

Lou, J. F., Lei, S. Y., Zhu, J. Q., and Wu, X. F. (2015). Chromosome Karyotype of *Pseudosciaena Crocea* Daiqu Population. *Chin. J. Zoolog.* 50 (1), 148–152. doi:10.13859/j.cjz.201501019

Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: Integrated Software for Genetic Linkage Map Construction and Quantitative Trait Locus Mapping in Biparental Populations. *Crop J.* 3, 269–283. doi:10.1016/j.cj.2015.01.001

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

Miao, L., Li, M. Y., Chen, J., and Zhang, H. (2014). Breeding of Fast Growth and Low Temperature Tolerance of New Variety Donghai No.1 Large Yellow Croaker (*Pseudosciaena Crocea*). *J. Agric. Biotechnol.* 22 (10), 1314–1320. doi:10.3969/j.issn.1674-7968.2014.10.015

Moen, T., Baranski, M., Sonesson, A. K., and Kjøglum, S. (2009). Confirmation and fine-mapping of a Major QTL for Resistance to Infectious Pancreatic Necrosis in Atlantic salmon (*Salmo salar*): Population-Level Associations between Markers and Trait. *BMC Genomics* 10, 368. doi:10.1186/1471-2164-10-368

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

Ning, Y., Liu, X., Wang, Z. Y., Guo, W., Li, Y., and Xie, F. (2007). A Genetic Map of Large Yellow Croaker Pseudosciaena Crocea. *Aquaculture* 264, 16–26. doi:10.1016/j.aquaculture.2006.12.042

Norris, A. (2017). Application of Genomics in salmon Aquaculture Breeding Programs by Ashie Norris. *Mar. Genomics* 36, 13–15. doi:10.1016/j.margen.2017.11.013

Ostberg, C. O., Hauser, L., Pritchard, V. L., Garza, J. C., and Naish, K. A. (2013). Chromosome Rearrangements, Recombination Suppression, and Limited Segregation Distortion in Hybrids between Yellowstone Cutthroat trout (*Oncorhynchus Clarkii Bouvieri*) and Rainbow trout (*O. mykiss*). *BMC Genomics* 14, 570. doi:10.1186/1471-2164-14-570

Palaiokostas, C., Bekaert, M., Khan, M. G. Q., Taggart, J. B., Gharbi, K., McAndrew, B. J., et al. (2013). Mapping and Validation of the Major Sex-Determining Region in Nile tilapia (*Oreochromis niloticus* L.) Using RAD Sequencing. *PLoS One* 8 (9), e68389–1371. doi:10.1371/journal.pone.0068389

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

Rastas, P., Paulin, L., Hanski, I., Lehtonen, R., and Auvinen, P. (2013). Lep-MAP: Fast and Accurate Linkage Map Construction for Large SNP Datasets. *Bioinformatics* 29, 3128–3134. doi:10.1093/bioinformatics/btt563

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., and Houston, R. (2017). Applications of Genotyping by Sequencing in Aquaculture Breeding and Genetics. *Rev. Aquacult* 10, 670–682. doi:10.1111/raq.12193

Sardell, J. M., Cheng, C., Dagilis, A. J., Ishikawa, A., Kitano, J., Peichel, C. L., et al. (2018). Sex Differences in Recombination in Sticklebacks. *G3 Genes|Genomes|Genetics* 8 (6), 1971–1983. doi:10.1534/g3.118.200166

Sardell, J. M., and Kirkpatrick, M. (2020). Sex Differences in the Recombination Landscape. *The Am. Naturalist* 195 (2), 361–379. doi:10.1086/704943

Siberchicot, A., Bessy, A., Guéguen, L., and Marais, G. A. (2017). MareyMap Online: a User-Friendly Web Application and Database Service for Estimating Recombination Rates Using Physical and Genetic Maps. *Genome Biol. Evol.* 9 (10), 2506–2509. doi:10.1093/gbe/evx178

Sonah, H., O'Donoughue, L., Cober, E., Rajcan, I., and Belzile, F. (2015). Identification of Loci Governing Eight Agronomic Traits Using a GBS-GWAS Approach and Validation by QTL Mapping in Soya Bean. *Plant Biotechnol. J.* 13 (2), 211–221. doi:10.1111/pbi.12249

Sun, S., and Xu, J. (2007). Genetic Analyses of a Hybrid Cross between Serotypes A and D Strains of the Human Pathogenic Fungus Cryptococcus Neoformans. *Genetics* 177 (3), 1475–1486. doi:10.1534/genetics.107.078923

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Gharbi, K., Stear, M. J., et al. (2015). Genome Wide Association and Genomic Prediction for Growth Traits in Juvenile Farmed Atlantic salmon Using a High Density SNP Array. *BMC Genomics* 16, 969. doi:10.1186/s12864-015-2117-9

Tsigenopoulos, C. S., Louro, B., Chatziplis, D., Lagnel, J., Vogiatzi, E., Loukovitis, D., et al. (2014). Second Generation Genetic Linkage Map for the Gilthead Sea Bream *Sparus Aurata* L. *Mar. Genomics* 18, 77–82. doi:10.1016/j.margen.2014.09.008

Turner, S. D. (2014). QQMAN: An R Package for Visualizing GWAS Results Using Q-Q and manhattan Plots. *bioRxiv*, 005165. doi:10.1101/005165

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

Wang, Z.-Y., and Cai, M.-Y. (2018). Artificial Gynogenesis and Sex Control in Large Yellow Croaker. *Sex Control Aquac. Artif. gynogenesis sex Control large yellow croaker* 11, 751–773. doi:10.1002/9781119127291.ch39

Xiao, S., Wang, P., Zhang, Y., Fang, L., Liu, Y., Li, J.-T., et al. (2015). Gene Map of Large Yellow Croaker (*Larimichthys Crocea*) Provides Insights into Teleost Genome Evolution and Conserved Regions Associated with Growth. *Sci. Rep.* 5, 18661. doi:10.1038/srep18661

Xu, D., Molina, W. F., Yano, C. F., Zhang, Y., de Oliveira, E. A., Lou, B., et al. (2017). Comparative Cytogenetics in Three Sciaenid Species (Teleostei, Perciformes): Evidence of Interspecific Chromosomal Diversification. *Mol. Cytogenet.* 10, 37. doi:10.1186/s13039-017-0338-0

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

Ye, H., Liu, Y., Liu, X., Wang, X., and Wang, Z. (2014). Genetic Mapping and QTL Analysis of Growth Traits in the Large Yellow Croaker Larimichthys Crocea. *Mar. Biotechnol.* 16, 729–738. doi:10.1007/s10126-014-9590-z

Yu, X. J., Xu, L. J., and Wu, F. X. (2020a). *China Fishery Statistical Yearbook*. Beijing, China: China Agriculture Press, 22–24.

Yu, X., Wu, C., and Gjøen, H. M. (2017a). Artificial Fertilization and Generating Families for a Selective Breeding Programme of Large Yellow Croaker (*Larimichthys Crocea*). *Int. Aquat. Res.* 9, 161–167. doi:10.1007/s40071-017-0164-3

Yu, X. X., Ådnøy, T., Lv, Z. M., Wu, C., and Gjøen, H. M. (2020b). Phenotypic and Genetic Parameter Estimation for Growth Traits in Juvenile Large Yellow Croaker (*Larimichthys Crocea*). *Fish. Aqua J.* 11, 274. doi:10.35248/2150-3508.20.11.274

Yu, Y., Ouyang, Y., and Yao, W. (2017b). shinyCircos: an R/Shiny Application for Interactive Creation of Circos Plot. *Bioinformatics* 34, 1229–1231. doi:10.1093/bioinformatics/btx763

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.2018.006910.3389/fgene.2018.00693

Zhou, Z., Han, K., Wu, Y., Bai, H., Ke, Q., Pu, F., et al. (2019). Genome-wide Association Study of Growth and Body-Shape-Related Traits in Large Yellow Croaker (*Larimichthys Crocea*) Using ddRAD Sequencing. *Mar. Biotechnol.* 21, 655–670. doi:10.1007/s10126-019-09910-0

Conflict of Interest: Author RJ is currently employed by GenoMar Genetics AS.

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

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 in Genetics

ORIGINAL RESEARCH
published: 08 February 2022
doi: 10.3389/fgene.2022.755004

Screening and Verification of Molecular Markers and Genes Related to Salt-Alkali Tolerance in *Portunus trituberculatus*

Wen Zhang[1,3], *Xiao Yan Zhao*[1], *Jie Wu*[1], *Ling Jin*[1], *Jianjian Lv*[1,2], *Baoquan Gao*[1,2] *and Ping Liu*[1,2]*

[1]Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, P.R.China, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China, [2]Function Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, [3]College of marine technology and environment, Dalian Ocean University, Dalian, China

Salt-alkali tolerance is one of the important breeding traits of *Portunus trituberculatus*. Identification of molecular markers linked to salt-alkali tolerance is prerequisite to develop such molecular marker-assisted breeding. In this study, Bulked Segregant Analysis (BSA) was used to screen molecular markers associated with salt-alkali tolerance trait in *P. trituberculatus*. Two DNA mixing pools with significant difference in salt-alkali tolerance were prepared and 94.83G of high-quality sequencing data was obtained. 855 SNPs and 1051 Indels were firstly selected as candidate markers by BSA analysis, out of which, 20 markers were further selected via △index value (close to 0 or 1) and eight of those were successfully verified. In addition, based on the located information of the markers in genome, eight candidate genes related to salt-alkali tolerance were anchored including ubiquitin-conjugating enzyme, aspartate–tRNA ligase, vesicle-trafficking protein, and so on. qPCR results showed that the expression patterns of all these genes changed significantly after salt-alkali stress, suggesting that they play certain roles in salt-alkali adaptation. Our results will provide applicable markers for molecular marker-assisted breeding and help to clarify the mechanisms of salt-alkali adaptation of *P. trituberculatus*.

Keywords: BSA, INDEL, Portunus trituberculatus, salt-alkali, SNP

OPEN ACCESS

Edited by:
Li Zhou,
Institute of Hydrobiology (CAS), China

Reviewed by:
Changkao Mu,
Ningbo University, China
Ce Shi,
Ningbo University, China

***Correspondence:**
Ping Liu
liuping@ysfri.ac.cn

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

Received: *07 August 2021*
Accepted: *12 January 2022*
Published: *08 February 2022*

Citation:
Zhang W, Zhao XY, Wu J, Jin L, Lv J, Gao B and Liu P (2022) Screening and Verification of Molecular Markers and Genes Related to Salt-Alkali Tolerance in Portunus trituberculatus. *Front. Genet. 13:755004. doi: 10.3389/fgene.2022.755004*

INTRODUCTION

Portunus trituberculatus (Juan and Gaoli, 2021; Ye and Fangmin, 2021), which belongs to Crustacea, Decapoda, Portunus family, commonly known as Portunus, is an important breeding species along the coast of China (Wu, 2014). Saline-alkali waters are a form of worldwide low-yield water resources, and there are about 690 million mu of low-lying saline-alkali waters in inland China alone. However, due to its high salinity, high alkalinity, high pH, and complex ion composition (Li, 2020), common aquatic animals cannot survive and reproduce in this environment suitably. The use of this type of water resource is thus, greatly hindered. High salt-alkali stimulation will reduce the feeding, metamorphosis, and survival rates of juvenile crabs (Hu et al., 2014). Therefore, the salt-alkali tolerance trait is one of the important breeding traits of *P. trituberculatus*.

TABLE 1 | Primers used in this study.

Primer ID	Primer sequence (5′-3′)
S1-F	AGGTCTCCCTCACAAACGC
S1-R	TGTGGATAGGAAAGGGTGAA
S2-F	CCATTACTTACTCTCACCATTTCAG
S2-R	CAACCTTGACGGAAACCATAC
S3-F	CTCACTTACCTGTCGTCACCTG
S3-R	GCACGCAGGTACTGAACATTT
S4-F	ACCACACCTGCCTAATCTACC
S4-R	CAAGATTCCAGTGTTTCTGTGAG
S5-F	AGTCACTATGAATGGCAAATATCTA
S5-R	CGGTTTGTAACTCTCGGGGT
I4-F	CTCTCCGCCAGCCCGCCATTAATGC
I4-R	TTACTTTCCATCCATCAGCC
I8-F	TGGGTGTTGTCAATCTGTGATGGCT
I8-R	CTGACTGACGAGACGACTGG
I9-F	TCGTCAGTCATCTTTCTTCTCTC
I9-R	TGTTTGGTTATTGGCGTTG

At present, research on salt-alkali tolerance in aquatic animals mainly focus on fishes, especially on species with natural resistance to salt-alkali, such as *Oncorhynchus clarkihenshawi* (Wilkie and Wright, 1994), and *Chalcalburnus tarichi* (Arabacı and Sarı, 2004). Studies of molecular physiological mechanisms have been carried out on *Oreochromis niloticus* (Biao and Li, 2012; Zhao and Wu, 2016) and *Oryzias latipes* (Yao and Wang, 2012). Some important ion transport genes including Na^+ -K+-ATPase, *slc4a1*, *slc4a4*, and *slc26a6*, were cloned and proved to play an important role in salt-alkali tolerance (Whittamore, 2012; Wood and Bergman, 2012). Compared with fish, less studies documented salt-alkali stress in aquatic crustaceans (Fu, 2002; Fu-yi et al., 2005; Li et al., 2021). Out of those, most studies are similar to the work in fish, focusing on expression of ion transport-related genes (Na^+/K^+-ATPase, HSP, proPO) (Shen et al., 2014; Xu et al., 2020) and the analysis of enzyme activity (ACP, AKP, SOD) (Liu, 2016; Xu et al., 2020) after salt-alkali stress. However, just a few cases were reported on whether crustaceans have special salt-alkali tolerance genes.

Molecular markers are prerequisite for marker-assisted breeding. Due to the important economic value of *P. trituberculatus*, some growth-, sex-, and immunity-related markers have been discovered and verified based on quantitative trait locus (QTL) and association analysis (Lv et al., 2017; Lv et al., 2018; Zhou and Zhao, 2019). Recently, in terms of environmental adaptation, some salinity-related molecular markers have also been reported (Lv et al., 2019). However, there were no reports on the development of salt-alkali tolerance molecular markers in crabs.

In this study, we explored the salt-alkali tolerance molecular markers in *P. trituberculatus* by the Bulked Segregant Analysis (BSA) strategy. In addition, based on the location information of the salt-alkali tolerance molecular markers on genome, we tried to anchor genes related to salt-alkali adaptation in crab. Our results will provide applicable markers for molecular marker-assisted breeding and help to clarify the mechanism of salt-alkali adaptation of *P. trituberculatus*.

MATERIALS AND METHODS

Experimental Material

The crab used in this study was cultured in Changyi Haifeng Aquaculture Co., Ltd., Weifang City, Shandong Province, China. Six hundred healthy crabs were selected and randomly divided into three groups (experimental, control, and sample), and the average weight was 55.28 g. During the temporary rearing (holding for 7 days), the water temperature was maintained at 22 ± 1 °C and pH 8.2 ± 0.5. Seawater of the experimental group and sample group was adjusted to 16.7 mmol/L by adding carbonate, and for the control group, normal seawater was used. Then, salt-alkali stress was observed every 2 h and dead individuals were collected in time. The experiment lasted for 72 h, until the survival rate of the experimental group was 5%, while no death was recorded in the control group. The first 20 individuals that died were from the saline-alkali-sensitive group (S) and the last 20 individuals that survived were from the saline-alkali-tolerant group (T). Muscle tissues from the crabs were collected and stored in liquid nitrogen. In the sampling group, gill tissues were sampled at 0, 6, 12, 24, 48, and 72 h after stress, and stored in liquid nitrogen for cDNA for qPCR detection analysis.

Library Construction and Sequencing

DNA degradation and contamination was monitored on 1% agarose gels. DNA purity was checked using NanoPhotometer® spectrophotometer (IMPLEN, CA, United States). DNA concentration was measured using Qubit® DNA Assay Kit in Qubit® 2.0 Fluorometer (LifeTechnologies, CA, United States). DNA sample was fragmented by sonication to a size of 350 bp, which were end polished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. PCR products were purified (AMPure XP system) and libraries were analyzed for size distribution. These libraries were sequenced by using Illumina HiSeq platform and 150 bp paired-end reads were generated.

TABLE 2 | Summary of sequencing data quality.

Group	Raw data (bp)	Valid data after filtering (bp)	Efficient (%)	Error rate (%)	Q20 (%)	Q30 (%)	GC content%)
S	45,854,625,300	45,753,784,500	99.78	0.04	93.81	87.06	40.62
T	49,159,475,100	49,083,786,900	99.85	0.04	93.92	87.31	40.86

TABLE 3 | Position information of candidate markers in the genome.

Category	Number of SNPs	Number of Indels
Upstream	12	27
Exonic-Stop gain	0	0
Exonic-Stop loss	0	0
Exonic-Synonymous	7	0
Exonic-Non-synonymous	2	0
Intronic	164	231
Splicing	0	0
Downstream	10	20
Upstream/Downstream	1	1
Intergenic	649	771
Ts	558	
Tv	297	
Ts/tv	1879	
Total	855	1051

BSA Analysis

Burrows-Wheeler Aligner (Heng and Richard, 2009) was used to align the clean reads of each sample against the reference genome. Alignment files were converted to BAM files using SAMtools software (Li et al., 2009). Variant calling was performed for all samples using the Unified Genotyper function in GATK3.3 (McKenna and Hanna, 2010) software. SNP was selected by using the Variant Filtration parameter in GATK and InDel was filtered by using the Variant Filtration parameter.

The reads depth information for the SNPs/InDels above in the offspring pools was gained to calculate the SNP/InDel index (Takagi and Abe, 2013). We filtered out those points whose SNP/InDel index in both pools were less than 0.3 and depth less than 7. The difference of the SNP/InDel index of two pools was

TABLE 4 | Genotyping information statistics.

Marker ID	Marker type	Genotype	Proportion of susceptible group	Proportion of tolerant group	*p*
S1	A/G	GG:AG:AA	18:2:0	4:10:6	0.000
S2	A/G	GG:AG:AA	12:6:2	20:0:0	0.007
S3	A/G	GG:AG:AA	15:1:0	6:10:0	0.001
S4	T/C	CC:CT:TT	11:7:2	2:11:7	0.007
S5	A/G	GG:AG:AA	0:3:17	12:0:8	0.000
I4	Indel	In:Del:Indel	0:15:5	13:5:2	0.000
I8	Indel	In:Del:Indel	5:2:13	18:0:2	0.000
I9	Indel	In:Del:Indel	7:4:9	0:16:4	0.000

TABLE 5 | Annotation information of salt-alkali tolerance candidate genes.

Gene ID	Category	△index	Annotation
Contig 574.9	Intergenic	−0.92	NR:Ubiquitin-conjugating enzyme E2 O [Zootermopsis nevadensis] KEGG:ubiquitin-conjugating enzyme E2-230k, putative (EC:6.3.2.19) K10581 ubiquitin-conjugating enzyme E2 O [EC:2.3.2.24] (A) SwissProt:E2/E3 hybrid ubiquitin-protein ligase UBE2O OS = Mus musculus GN = Ube2o PE = 1 SV = 3; IPR_id:NULL; IPR_Anno:NULL
Contig 11.37	Intergenic	0.74	NR:hypothetical protein X975_00935, partial [Stegodyphus mimosarum] KEGG:putative ZDHHC-type palmitoyltransferase 6; K20032 palmitoyltransferase ZDHHC13/17 [EC:2.3.1.225] (A); SwissProt:Espin OS = Mus musculus GN = Espn PE = 1 SV = 2 IPR_id:NULL; IPR_Anno:NULL
Contig 272.2	Intergenic	0.71	NR:PREDICTED: aspartate--tRNA ligase, mitochondrial [Tribolium castaneum] KEGG:kinesin-6A; kinesin 6A; K17387 kinesin family member 23 (A) SwissProt:Aspartate--tRNA ligase, mitochondrial OS = Rattus norvegicus GN = Dars2 PE = 1 SV = 1 IPR_id:IPR008126; IPR_Anno:Outer membrane adhesion, Yersinia
Contig 159.21	Intergenic	0.78	NR:PREDICTED: gamma-tubulin complex component 6 isoform X2 [Pogonomyrmex barbatus] KEGG:TUBGCP6; tubulin, gamma complex associated protein 6 K16573 gamma-tubulin complex component 6 (A) SwissProt:Gamma-tubulin complex component 6 OS = Homo sapiens GN = TUBGCP6 PE = 1 SV = 3; IPR_id:IPR007259; IPR_Anno:Spc97/Spc98
Contig 105.16	Intergenic	0.79	NR:PREDICTED: uncharacterized protein LOC105387196 [Plutella xylostella] KEGG: ; SwissProt: ; IPR_id:NULL; IPR_Anno:NULL
Contig 48.16	Intergenic	−0.77	NR: ; KEGG: ; SwissProt:Transcription termination factor 1 OS = Homo sapiens GN = TTF1 PE = 1 SV = 3; IPR_id:IPR005448; IPR_Anno:Voltage-dependent calcium channel, P/Q-type, alpha-1 subunit
Contig 2378.3	Intergenic	0.75	NR:PREDICTED: vesicle-trafficking protein SEC22b-B [Fopius arisanus] KEGG:vesicle-trafficking protein SEC22b-B; K08517 vesicle transport protein SEC22 (A) SwissProt:Vesicle-trafficking protein SEC22b-B OS = Danio rerio GN = sec22bb PE = 2 SV = 1; IPR_id:IPR002255; IPR_Anno:Flavin monooxygenase (FMO) 3
Contig 370.5	Intergenic	−0.72	NR:PREDICTED: protein SMG5 [Acromyrmex echinatior] KEGG:protein SMG5; K11125 protein SMG5 (A) SwissProt:Protein SMG5 OS = Homo sapiens GN = SMG5 PE = 1 SV = 3 IPR_id:IPR018834; IPR_Anno:DNA/RNA-binding domain, Est1-type

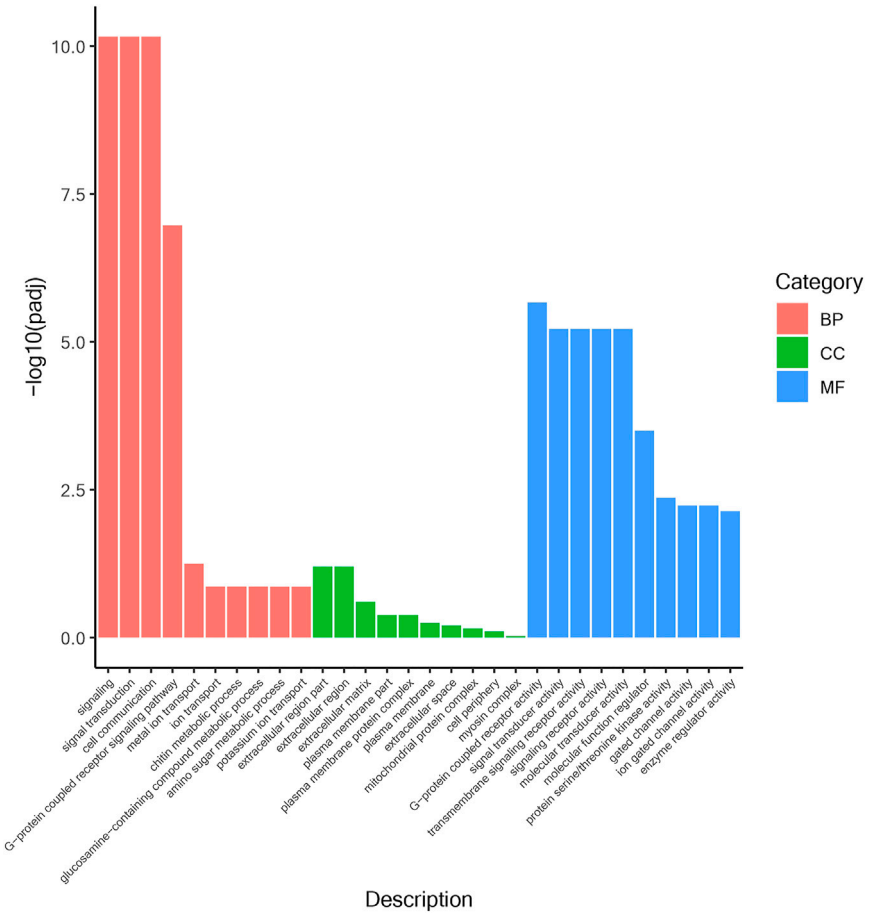

FIGURE 1 | GO enrichment analysis of salt-alkali tolerance candidate genes.

calculated as the △ SNP/InDel index. The markers with an absolute value of △index between 0.69 and 1 were screened out as candidate molecular markers.

Validation of Molecular Markers Related to Salt-Alkali Tolerance

SNPs and Indels conducted two rounds of verification (**Table 1**). The first round used a mixed template made by mixing the extracted DNA individual templates for PCR amplification and preliminary screening. The second round used individual DNA templates for amplification and checked by sequencing, or 2% agarose electrophoresis. Statistics of the genotype of each individual were based on the results of sequencing and electrophoresis. SPSS 17.0 software was used for one-way analysis of variance (ANOVA), and $p < 0.05$ was considered to be significantly related to salt-alkali tolerance.

qPCR

The volume of the qPCR reaction system was 10 μL, containing 5 μL of 2 × SYBR Master Mix (RR420A, Takara Bio, Japan), 0.2 μL of each primer and ROX Reference Dye II, 1 μL of cDNA template, and 3.4 μL of rnase Free dH2O. qPCR was carried out in a FAST-7500 system (ABI-7500, ThermoFisher, Singapore) as follows: 95°C or 30 s, and 40 cycles of 95°C for 5 s and 60°C for 34 s. Results from qPCR were analyzed by SPSS software's ANOVA for group differences, and Origin 9.1 was used for mapping.

RESULTS

Screening of Candidate Molecular Markers

In this study, raw data of 45,854,625,300 and 49,159,475,100 were obtained from the populations with significant differences to salt-alkali tolerance (**Table 2**). The sequencing depth is

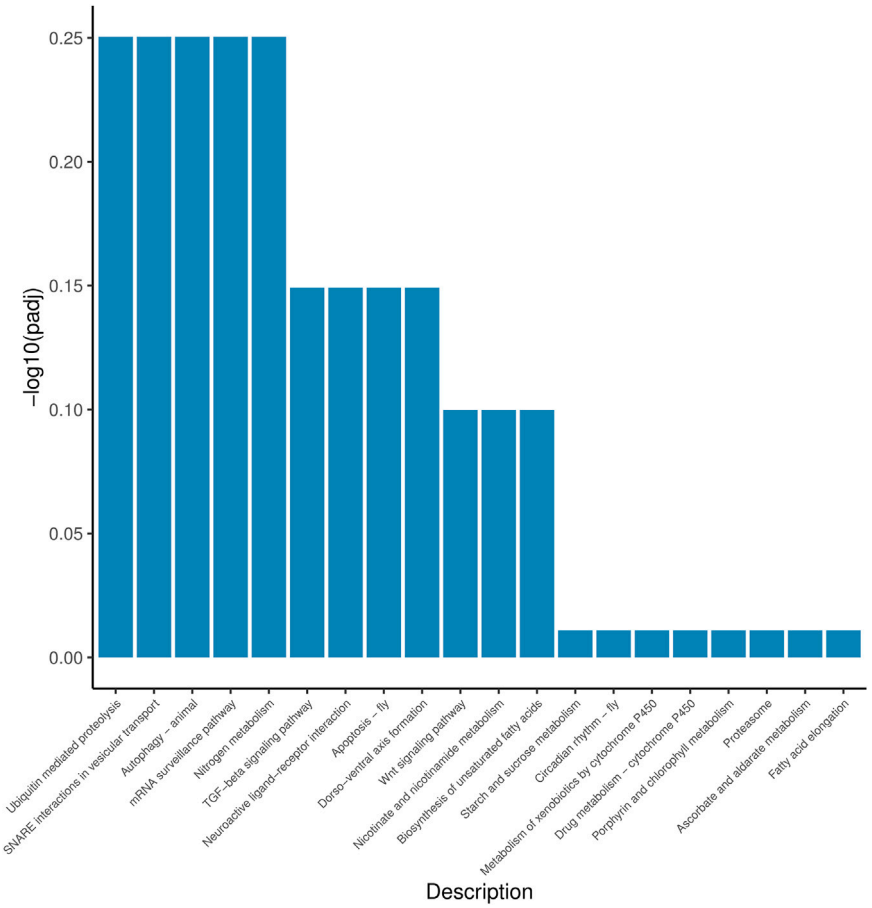

FIGURE 2 | KEGG enrichment analysis of salt-alkali tolerance candidate genes.

44.65X and 48.31X. A total of 855 SNPs (**Supplementary Table S1**) and 1051 Indels candidate polymorphic sex sites were selected with 95% confidence level (**Supplementary Table S2**). Among them, 395 were in the intron region and 1,420 in the intergenic region (**Table 3**).

Validation of Molecular Markers

A total of ten SNPs and ten Indels were screened out based on △index value (close to 0 or 1). Using the mixed template for PCR amplification, a total of 20 sites were successfully amplified by the target PCR product. Then, 40 individuals were typed for the successfully verified markers in the mixed template by PCR product sequencing and gel agarose electrophoresis. The association analysis test was performed by SPSS, and eight of the markers showed significant correlation with salt-alkali tolerance ($p < 0.05$) (**Table 4**), including five SNPs and three Indels (**Table 5**).

Enrichment Analysis of Candidate Genes

According to the position information of the candidate marker in the genome, a total of 2171 genes were located, which were used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. The GO results (**Figure 1**) showed that 30 processes were significantly enriched, of which "signaling", "extracellular region part", and "G-protein coupled receptor activity" were the main processes with the largest number of genes in "Biological Process", "Cellular Component" and "Molecular Function", respectively. Besides, a total of 20 KEGG pathways (**Figure 2**) were enriched, among which "Ubiquitin mediated proteolysis", "SNARE interactions in vesicular transport", "Autophagy", "mRNA surveillance pathway" and "Nitrogen metabolism" were the most enriched pathways ($p < 0.05$).

Expression Analysis of Candidate Genes

In the light of gene annotation and enrichment analysis results, eight genes related to proteolysis, autophagy, nitrogen metabolism, and vesicle transport were screened. qPCR results showed that the expression patterns of these genes changed significantly after salt-alkali stress (**Figure 3**), suggesting that they play a role in the adaptation to salt-

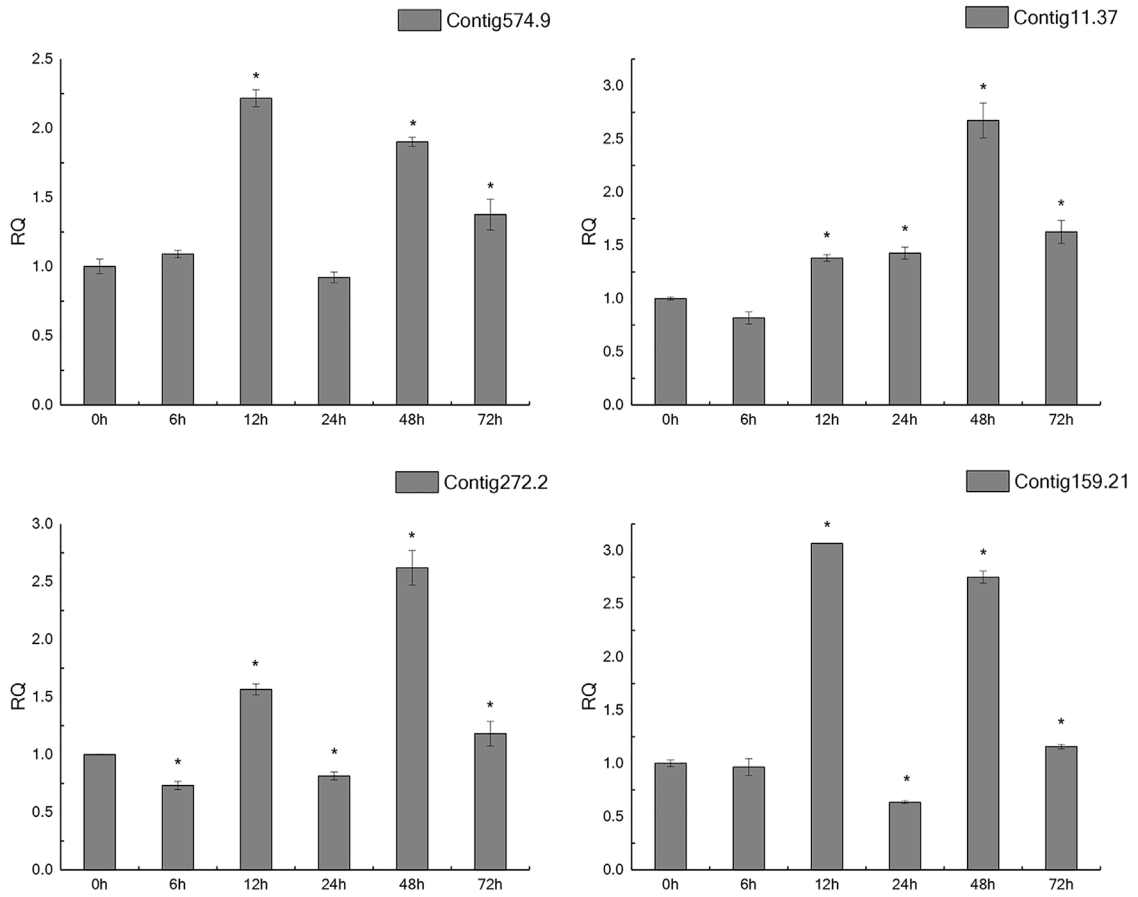

FIGURE 3 | qPCR results of salt-tolerant-related genes.

alkali. The results show that these genes can be divided into two types (**Figure 4**). One group was significantly up-regulated at 48 h and the other group of genes was significantly up-regulated at 12 and 48 h with a trend of fluctuating expression.

DISCUSSIONS

In this study, we successfully screened the salt-alkali tolerance markers in *P. trituberculatus,* using the BSA strategy for the first time, and then anchored several genes related to salt-alkali tolerance. The BSA method that was proposed by Michelmore and Paran, (1991) can lock the target trait genes or markers by comparing the degree of difference in the allele frequency at the polymorphic site between the two groups with markedly differentiated traits (June 2021; Rujia and Yiming, 2021; Zaiqing and Anmin, 2021). At present, research on trait-related markers of aquatic animals is mainly based on QTL and association analysis (Wang and Liu, 2017; Cao and Wang, 2021). QTL analysis relies on pedigree materials, and only excavated markers are applicable (Zhixue and Khorshed, 2021). Association analysis is often based on re-sequencing and the cost is relatively high (Hussain and Rivandi, 2007; Vermerris, 2011; Ali, 2013). Compared with the above two methods, BSA has the advantages of being able to use population materials and being cost effective and time efficient (Takagi Abe, 2013). Using the BSA strategy, trait-related markers have been successfully developed in plants (Cheng and Abass, 2020; Marc, 2020; Farhan and Qiutao, 2021; Hossein and Azin, 2021), proving it to be a feasible and economical method. However, its applicability in aquatic animals had not been reported in the past. In this study, 20 markers were selected based on BSA analysis, and eight salt-alkali tolerance related markers were finally developed efficiently. The results of the study show that BSA strategy is suitable for the mining of trait-related markers in aquatic crustaceans.

We noticed that most of the successfully verified SNPs and Indels are located on the intergenic region (81.82%). The intergenic region was once considered unimportant. However, in recent years, it has been found that these regions contain functionally important elements (Cai, 2016) and independent transcription of miRNAs with their own promoters (Hess et al., 2006). Intergenic regions may also contain unidentified functional elements, such as non-coding RNA (Zhong et al., 2014; Fariha and Changrui, 2019). It is currently believed that intergenic

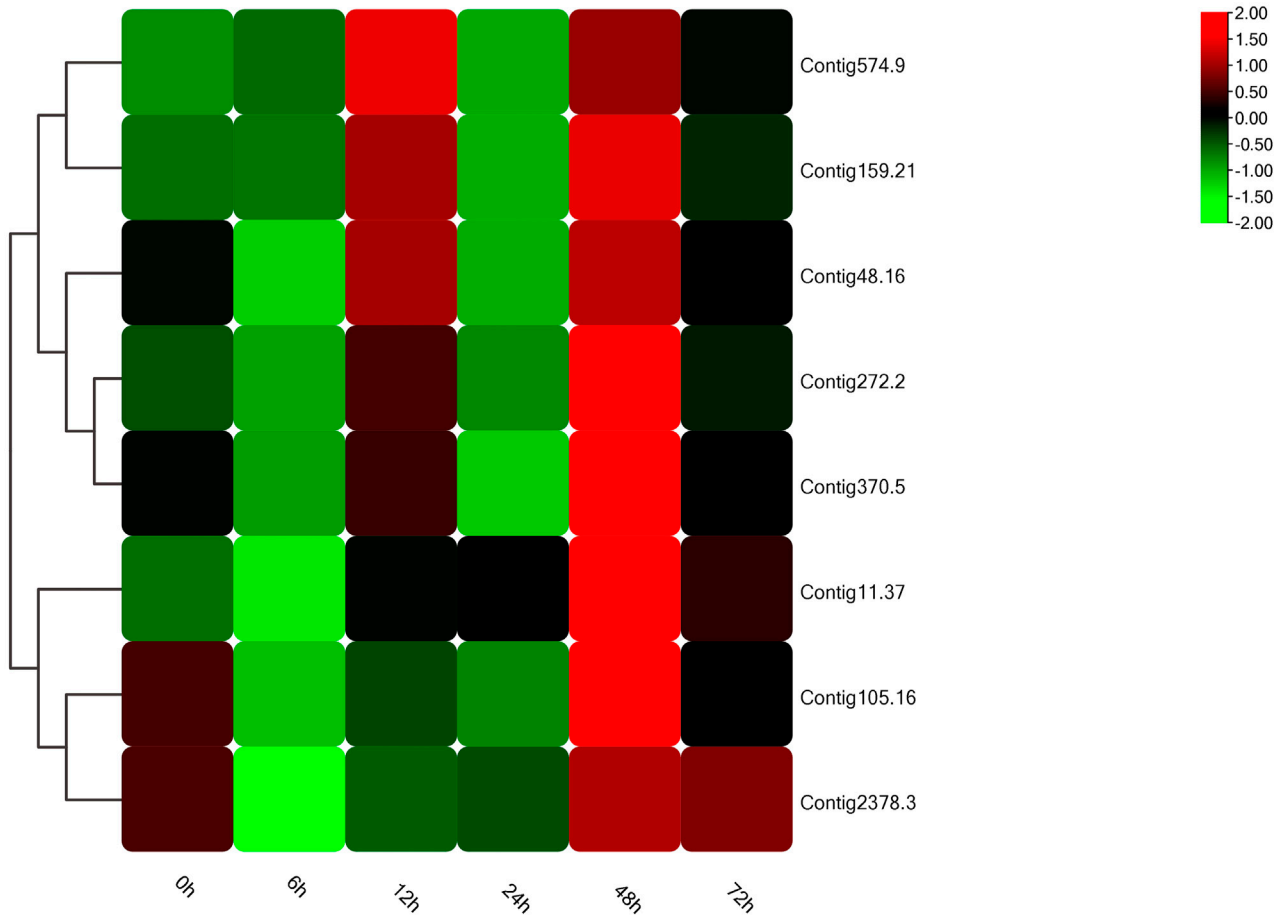

FIGURE 4 | Heat map of salt-tolerant-related genes.

regions can control gene expression and inhibition, and help RNA to identify genes with low conservation and diversity (Ingvarsson et al., 2016; Bakhtiarizadeh and Salehi, 2018). The fact that most of the saline-alkali markers are located in the intergenic region, indicates that some potential functional regions of the intergenic region may have important functions in the regulation of saline-alkali adaptation. However, the specific mechanisms are still unclear and necessitate further research.

Based on the location information of markers in the genome, we initially anchored 2171 potential saline-alkali adaptation genes, which were mainly enriched in the pathway including ubiquitin-mediated proteolysis, SNARE interactions in vesicle transport and autophagy. These results indicate that the mechanism of saline-alkali adaptation is complex and co-regulated by an equally complex regulatory network. It is worth noting that the salt-alkali-related ion transport genes such as Na^+-K^+-ATPase are not included here. This suggests that, although ion transport plays an important role in the process of saline-alkali adaptation, it does not play a prominent role in the differentiation of salt-alkali tolerance.

In addition, according to annotation information, 8 genes were screened from the five most enriched KEGG pathways for qPCR verification. The expression patterns of these eight genes changed significantly after salt-alkali stress, indicating that they may have certain functions in salt-alkali adaptation, or salt-alkali stress has affected their normal functions. We also found that most genes were up-regulated at 12 and 48 h, suggesting the critical time for salt-alkali adaptation.

In this study, we screened the salt-tolerant markers to prove the feasibility of the BSA strategy in aquatic research. The 8 successfully verified markers provide usable markers for subsequent molecular marker-assisted breeding. Furthermore, this experiment anchored some candidate genes related to salt-alkali tolerance. These results will help to systematically clarify the molecular mechanism of salt-alkali adaptation of *P. trituberculatus*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available in NCBI using accession number PRJNA755748.

AUTHOR CONTRIBUTIONS

WZ: conducted experiments, analyzed data and results, paper writing. XZ: conducted experiments. JW: collect the samples. LJ: idea for the project. JL: analyzed data and results, paper revising. BG: idea for the project. PL: paper revising.

FUNDING

This research was supported by the National Key RD Program of China (2018YFD0901304), Ministry of Agriculture of China and China Agriculture Research System (CARS-48), National Natural Science Foundation of China (42076116) and Projects of International Exchange and Cooperation in Agriculture, Ministry of Agriculture and Rural Affairs of China-Science, Technology and Innovation Cooperation in Aquaculture with Tropical Countries along the Belt and Road.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.755004/full#supplementary-material

REFERENCES

Ali, F. (2013). *Exploring the Genetic Basis of Northern Leaf Blight via Genome Wide Association Mapping in Maize*. Wuhan, China: Huazhong Agricultural University.

Arabacı, M., and Sarı, M. (2004). Induction of Ovulation in Endemic Pearl Mullet (Chalcalburnus Tarichi), Living in the Highly Alkaline Lake Van, Using GnRHa ([D-Ser(T Bu)6, Pro9-Net]-GnRH) Combined with Haloperidol. *Aquaculture.* 238 (1), 529–535. doi:10.1016/j.aquaculture.2004.05.004

Bahadur, K., and Tripathi, P. (1978). Investigation if Inorganic and Organic Nitrogenous Compounds on Microbial Nitrogen Fixation with Respect to the Fixation Cycle. *Zentralbl Bakteriol Naturwiss.* 133 (2), 154. doi:10.1016/s0323-6056(78)80013-5

Bakhtiarizadeh, M. R., Salehi, A., and Rivera, R. M. (2018). Genome-wide Identification and Analysis of A-To-I RNA Editing Events in Bovine by Transcriptome Sequencing. *PLOS ONE.* 13 (2), e0193316. doi:10.1371/journal.pone.0193316

Biao, Y., and Li, Z. (2012). miR-429 Regulation of Osmotic Stress Transcription Factor 1 (OSTF1) in tilapia during Osmotic Stress. *Biochem. biophysical Res. Commun.* 426 (3), 294–298. doi:10.1016/j.bbrc.2012.08.029

Cai, L. (2016). *Comprehensive Characterization of LincRNAs' Functions in Transcription Regulation through Long-Range Chromatin Interactions*. Wuhan, China: Huazhong Agricultural University.

Cao, j., and Wang, w. (2021). Construction of a High-Density Genetic Linkage Map and QTL Mapping for Gender of *Megalobrama amblycephala*. *J. Huazhong Agric. Univ.* 40 (02), 188–196. doi:10.13300/j.cnki.hnlkxb.2021.02.020

Cheng, Q., and Abass, A. M. (2020). Comparative Transcriptome Analysis Reveals the Regulatory Effects of Acetylcholine on Salt Tolerance of Nicotiana Benthamiana. *Phytochemistry.* 181, 112582. doi:10.1016/j.phytochem.2020.112582

De Carvalho, I. P. C., and Detmann, E. (2012). *In Vitro* and *In Situ* Activity of Carboxymethyl Cellulase and Glutamate Dehydrogenase According to Supplementation with Different Nitrogenous Compounds. *Revista Brasileira de Zootecnia.* 41 (3), 683. doi:10.1590/s1516-35982012000300031

Evans, D. H. (2010). A Brief History of the Study of Fish Osmoregulation: The Central Role of the Mt. Desert Island Biological Laboratory. *Front. Physiol.* 1, 13. doi:10.3389/fphys.2010.00013

Farhan, U., and Qiutao, X. (2021). Histone Deacetylase HDA710 Controls Salt Tolerance by Regulating ABA Signaling in rice. *J. Integr. Plant Biol.* 63 (3), 451–467. doi:10.1111/jipb.13042

Fariha, K., and Changrui, L. (2019). A Review on Native and Denaturing Purification Methods for Non-coding RNA (ncRNA). *J. Chromatogr. B, Anal. Tech. Biomed. Life Sci.* 1120, 71–79. doi:10.1016/j.jchromb.2019.04.034

Fu, L. (2002). *Effcts of Environmental Factors on the Osmoregulation of Eriocheir Sinensis*. Qingdao, China: Ocean University of China.

Fu-yi, Y., Xiu-jun, L., Wang, Z-c., Zhao, C-s., and Sun, L-m. (2005). Possibility Test of Transplanting Prawn in Carbonate Saline-Alkali Waters. *J. offilin Agric. Univ.* 27 (5), 559–564. doi:10.1360/biodiv.050121

Hassan, N. M., and Fadwa, S. (2021). Dysregulation of CCAAT/enhancer Binding Protein-Alpha (CEBPA) Expression in the Bone Marrow of Acute Myeloid Leukemia Patients. *Egypt. J. Med. Hum. Genet.* 22 (1), 30. doi:10.1186/s43042-021-00154-z

Hess, N. K., Singer, P. A., Trinh, K., Nikkhoy, M., and Bernstein, S. I. (2006). Transcriptional Regulation of the *Drosophila melanogaster* Muscle Myosin Heavy-Chain Gene. *Gene Expr. Patterns.* 7 (4), 413–422. doi:10.1016/j.modgep.2006.11.007

Hossein, S., and Azin, T. (2021). Crushed maize Seeds Enhance Soil Biological Activity and Salt Tolerance in Caper (Capparis Spinosa L.). *Ind. Crops Prod.* 160, 113103. doi:10.1016/j.indcrop.2020.113103

Hu, Z., Li, Z., Yu, J., Tong, C., Lin, Y., Guo, X., et al. (2014). Association Analysis Identifies New Risk Loci for Non-obstructive Azoospermia in Chinese Menfluences of Different Salinities on Growth and Survival Rate of Portunus Trituberculatus. *Nat. Commun.* 5 (10), 3857–3866. doi:10.1038/ncomms4857

Hussain, S. S., and Rivandi, A. (2007). Molecular Breeding for Drought Tolerance in Plants:Wheat Perspective. *Proc. Pak. Acad. Sci.*, 31–58.

Ingvarsson, P. K., Hvidsten, T. R., and Street, N. R. (2016). Towards Integration of Population and Comparative Genomics in forest Trees. *New Phytol.* 212 (2), 338–344. doi:10.1111/nph.14153

Juan, D., and Gaoli, Z. (2021). High-intensity Light of Full-Spectrum LED Promotes Survival Rate but Not Development of the Larval Swimming Crab Portunus Trituberculatus. *Aquacultural Eng.* 93, 102158. doi:10.1016/j.aquaeng.2021.102158

Jun, L. J. (2021). Comparative Transcriptomics and RNA-Seq-Based Bulked Segregant Analysis Reveals Genomic Basis Underlying Cronartium Ribicola Vcr2 Virulence& 13. *Front. Microbiol.* 12, 602812. doi:10.3389/fmicb.2021.602812

Kai, W., and Mingyao, L. (2010). ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data. *Narnia.* 38 (16).

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics.* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Wysoker, B. A., Fennell, T., Ruan, J., Homer, N., Marth, G., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics.* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Li, L. (2020). Green Culture Technology in Saline-Alkali Water. *Agric. knowledge.*, S964.

Li, M., Li, J., Shi, K., He, Y., Gao, B., Liu, P., et al. (2021). Estimation of Heritability and Genetic Correlation of saline-alkali Tolerance in Exopalaemon Carinicauda. *Prog. Fish. Sci.* 42 (1), 117–123. doi:10.19663/j.issn2095-9869.20200221001

Liu, F. (2016). *Effects of Carbonate Alkalinity Stresses on the Growth and Reproduction of Exopalaemon Carinicauda*. Shanghai, China: Shanghai Ocean University.

Lv, J., Sun, D., Huan, P., Song, L., Liu, P., and Li, J. (2018). QTL Mapping and Marker Identification for Sex-Determining: Indicating XY Sex Determination System in the Swimming Crab (Portunus Trituberculatus). *Front. Genet.* 9, 337. doi:10.3389/fgene.2018.00337

Lv, J., Sun, D., Yan, D., Ti, X., Liu, P., and Li, J. (2019). Quantitative Trait Loci Mapping and Marker Identification for Low Salinity Tolerance Trait in the Swimming Crab (Portunus Trituberculatus). *Front. Genet.* 10, 1193. doi:10.3389/fgene.2019.01193

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

Marc, S. (2020). How a Mangrove Tree Can Help to Improve the Salt Tolerance of Arabidopsis and Rice. *Plant Physiol.* 184 (4), 1630–1632. doi:10.1104/pp.20.01370

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 (9), 1297–1303. doi:10.1101/gr.107524.110

Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of Markers Linked to Disease-Resistance Genes by Bulked Segregant Analysis: A Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations. *Proc. Natl. Acad. Sci. U S A.* 88 (21), 9828–9832. doi:10.1073/pnas.88.21.9828

Nuo, S., and Zeya, Z. (2021). Generation of an Induced Pluripotent Stem Cell Line from a Congenital Microtia Patient with 4p16.1 Microduplication Involving the Long-Range Enhancer of HMX1. *Stem Cel Res.*

Rujia, Z., and Yiming, R. (2021). Mapping of Genetic Locus for Leaf Trichome Formation in Chinese Cabbage Based on Bulked Segregant Analysis. *Plants.* 10 (4), 771. doi:10.3390/plants10040771

Ryoji, K., and Yohko, K. (2021). Distinct Foxp3 Enhancer Elements Coordinate Development, Maintenance, and Function of Regulatory T Cells. *Immunity.*

Shen, A., Jiang, K., and Shen, X. (2014). Effects of Salinity on the mRNA Expression of Na^+/K^+-ATPase and HSP in the Ridgetail white Prawn, *Exopalaemon Carinicauda*. *Mar. Environ. Sci.* 33 (05), 693–698.

Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, C., et al. (2013). QTL-seq: Rapid Mapping of Quantitative Trait Loci in rice by Whole Genome Resequencing of DNA from Two Bulked Populations. *Plant J.* 74 (1), 174–183. doi:10.1111/tpj.12105

Vermerris, W. (2011). Survey of Genomics Approaches to Improve Bioenergy Traits in maize, Sorghum and Sugarcane. *J. Integr. Plant Biol.* 53 (02), 105–119. doi:10.1111/j.1744-7909.2010.01020.x

Wang, X., Liu, S., Jiang, C., Geng, X., Zhou, T., Li, N., et al. (2017). Multiple Across-Strain and Within-Strain QTLs Suggest Highly Complex Genetic Architecture for Hypoxia Tolerance in Channel Catfish. *Mol. Genet. Genomics.* 292 (1), 63–76. doi:10.1007/s00438-016-1256-2

Whittamore, J. M. (2012). Osmoregulation and Epithelial Water Transport: Lessons from the Intestine of marine Teleost Fish. *J. Comp. Physiol. B.* 182 (1), 1–39. doi:10.1007/s00360-011-0601-3

Wilkie, M. P., and Wright, P. A. (1994). The Physiological Adaptations of the Lahontan Cutthroat Trout (*Oncorhynchus clarki* Henshawi) Following Transfer from Well Water to the Highly Alkaline Waters of Pyramid Lake, Nevada (pH 9.4). *Physiol. Zoolog.* 67 (2), 355–380. doi:10.1086/physzool.67.2.30163853

Wood, C. M., Bergman, H. L., Bianchini, A., Laurent, P., Maina, J., Johannsson, O. E., et al. (2012). Transepithelial Potential in the Magadi tilapia, a Fish Living in Extreme Alkalinity. *J. Comp. Physiol. B.* 182 (2), 247–258. doi:10.1007/s00360-011-0614-y

Wu, X., Wang, Q., Lou, B., Liu, Z., and Cheng, Y. (2014). Effects of Fattening Period on Ovarian Development and Nutritional Quality of Female Swimming Crab (Portunus Trituberculatus). *Journal Offisheries China.* 38 (02), 170–182. doi:10.3724/SP.J.1231.2014.48934

Xu, Z., Zhang, J., Zhang, W., Liang, J., Jian, L., and Li, J. (2020). Enzyme Activity and Gene Expressions of Immune-Related in Exopalaemon Carinicauda Following Long-Term Different Salinity Exposure. *Henan Fish.* 4 (01), 17–20.

Yao, Z. L., and Wang, H. (2012). Transcriptomic Profiles of Japanese Medaka (*Oryzias latipes*) in Response to Alkalinity Stress. *Genet. Mol. Res. : GMR* 11 (3), 2200. doi:10.4238/2012.June.15.2

Ye, Y., and Fangmin, X. (2021). Untargeted Lipidomics Reveals Metabolic Responses to Different Dietary N-3 PUFA in Juvenile Swimming Crab (Portunus Trituberculatus). *Food Chem.* 354, 129570. doi:10.1016/j.foodchem.2021.129570

Zaiqing, W., and Anmin, Y. (2021). Bulked Segregant Analysis Reveals Candidate Genes Responsible for dwarf Formation in Woody Oilseed Crop castor Bean. *Scientific Rep.* 11 (1), 6277. doi:10.1038/s41598-021-85644-1

Zhao, Y., Wu, J. W., Wang, Y., and Zhao, J. L. (2016). Role of miR-21 in Alkalinity Stress Tolerance in tilapia. *Biochem. Biophys. Res. Commun.* 471 (1), 26–33. doi:10.1016/j.bbrc.2016.02.007

Zhixue, D., and Khorshed, A. M. (2021). Mapping of a Major QTL Controlling Plant Height Using a High-Density Genetic Map and QTL-Seq Methods Based on Whole-Genome Resequencing in Brassica Napus. *G3: Genes, Genomes, Genet.* 11 (7), jkab118. doi:10.1093/g3journal/jkab118

Zhong, C., Andrews, J., and Zhang, S. (2014). Discovering Non-coding RNA Elements in Drosophila 3' Untranslated Regions. *Int. J. Bioinform Res. Appl.* 10 (4/5), 479–497. doi:10.1504/IJBRA.2014.062996

Zhou, S., and Zhao, J. (2019). Characterization, Subcellular Localization and Function Analysis of Myeloid Differentiation Factor 88 (Pt -MyD88) in Swimming Crab, Portunus Trituberculatus. *Fish Shellfish Immunol.* 95, 227. doi:10.1016/j.fsi.2019.10.036

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 in Genetics

ORIGINAL RESEARCH
published: 15 February 2022
doi: 10.3389/fgene.2022.852280

A High-Density Genetic Map and QTL Fine Mapping for Growth- and Sex-Related Traits in Red Swamp Crayfish (*Procambarus clarkii*)

Xin-Fen Guo[1,2], *Yu-Lin Zhou*[1,2,3], *Min Liu*[1,2], *Zhi Li*[1], *Li Zhou*[1,2], *Zhong-Wei Wang*[1,2]* and *Jian-Fang Gui*[1,2]

[1]State Key Laboratory of Freshwater Ecology and Biotechnology, Hubei Hongshan Laboratory, The Innovation Academy of Seed Design, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China, [2]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, [3]Key Laboratory of Ministry of Water Resources for Ecological Impacts of Hydraulic-Projects and Restoration of Aquatic Ecosystem, Institute of Hydroecology, Ministry of Water Resources, Chinese Academy of Sciences, Wuhan, China

Red swamp crayfish (*Procambarus clarkii*) is a commercially important species in global aquaculture and most successfully invasive freshwater shrimp in China. In order to determine the genetic basis of growth- and sex-related traits, a high-density genetic linkage map was constructed using 2b-RAD sequencing technology in a full-sib family. The consensus map contains 4,878 SNP markers assigned to 94 linkage groups (LGs) and spanned 6,157.737 cM with an average marker interval of 1.26 cM and 96.93% genome coverage. The quantitative trait locus (QTL) mapping for growth and sex traits was performed for the first time. QTL mapping uncovers 28 QTLs for growth-related traits in nine LGs, explaining 7.9–14.4% of the phenotypic variation, and identifies some potential candidate growth-related genes such as *mih*, *lamr*, *golgb1*, *nurf301*, and *tbcd1* within the QTL intervals. A single major locus for sex determination was revealed in LG20 that explains 59.3–63.7% of the phenotypic variations. Some candidate sex-related genes, such as *vps4bl*, *ssrf*, and *acot1*, were identified in the QTL intervals and found to be differentially expressed in the muscle tissues between the females and the males. Furthermore, the identified SNPs were revealed to be female heterozygotes, suggesting that red swamp crayfish might have the female heterogametic ZZ/ZW sex determination system. The present study provides a valuable resource for marker-assisted selection and genetic improvement and for further genetic and genomic research in red swamp crayfish.

Keywords: ***Procambarus clarkii*, genetic linkage map, QTL mapping, growth-related traits, sex**

OPEN ACCESS

Edited by:
Jun Hong Xia,
Sun Yat-sen University, China

Reviewed by:
Changxu Tian,
Guangdong Ocean University, China
Yubang Shen,
Shanghai Ocean University, China

***Correspondence:**
Zhong-Wei Wang
wzw0909@ihb.ac.cn

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

Received: *11 January 2022*
Accepted: *26 January 2022*
Published: *15 February 2022*

Citation:
Guo X-F, Zhou Y-L, Liu M, Li Z, Zhou L, Wang Z-W and Gui J-F (2022) A High-Density Genetic Map and QTL Fine Mapping for Growth- and Sex-Related Traits in Red Swamp Crayfish (Procambarus clarkii). Front. Genet. 13:852280. doi: 10.3389/fgene.2022.852280

INTRODUCTION

The red swamp crayfish (*Procambarus clarkii*) was native to the United States and Mexico and was introduced to Nanjing, China, from Japan in the 1930s (Hobbs et al., 1989; Yue et al., 2008). The crayfish has become widespread in almost all forms of freshwater such as lakes, rivers, and rice paddy fields all around China in the past few decades due to its high environmental adaptability and survivability (Shen et al., 2014). Furthermore, at present, red swamp crayfish has a great market

demand because of its delicious meat in China. In addition to its food value, it is also a good crustacean model organism in the studies of viral infections and environmental stress (Wu et al., 2012; Celi et al., 2013; Chen et al., 2013; Lin et al., 2013). Growth is an important trait that determines economic benefit in aquaculture. The traditional breeding model in crayfish, namely, catching and selling large individuals but maintaining small individuals as parents in rice fields, has resulted in a decrease of growth traits remarkably. Furthermore, sexual dimorphism was observed, and it showed that the males grow faster than the females, but the sex-determining mechanism of crayfish is still unclear. So, there is an urgent need to establish a precise breeding program to improve economic traits and promote the development of crayfish industry (Gui et al., 2022). Traditional genetic breeding programs mainly rely on phenotypic trait selection, which has a long breeding cycle and low selection efficiency (Zenger et al., 2017). Fast and effective breeding methods, such as marker-assisted selection (MAS), can use genetic information to select important traits, which provides convenience for genetic selection (Raina et al., 2020).

In recent years, the genetic linkage map construction and QTL mapping have become important marker-assisted selection breeding tools for their abilities to identify molecular markers or candidate genes, which are associated with traits of growth, sex, and disease resistance (Yu et al., 2015; Peng et al., 2016). In the early years, the molecular markers including AFLP, RFLP, and SSR were mainly used for construction of genetic linkage maps (Yue, 2014). However, the development of these markers is expensive and time consuming, resulting in fewer and low-density molecular markers, limiting their ability for fine mapping and localization of QTL for important traits. At present, single-nucleotide polymorphisms (SNPs) are the most abundant genetic markers in genome, which are stable, simple, and may be directly related to the phenotype (Baird et al., 2008; Shi et al., 2014). Therefore, SNPs gradually replaced the traditional molecular markers and became the best choice in the construction of the genetic linkage map (Fu et al., 2016; Peng et al., 2016; Liu et al., 2017; Feng et al., 2018; Zhang et al., 2019; Zhou et al., 2021). The development of next-generation sequencing technologies has facilitated the development of SNPs, and various genotyping methods have been developed to screen thousands of SNP markers rapidly and cheaply (Davey et al., 2011). Restriction site-associated DNA sequencing (RAD-seq) uses specific restriction enzymes to identify short fragments of DNA near the site to reduce the complexity of target genome samples (Baird et al., 2008; Davey and Blaxter, 2010). 2b-RAD-seq, a modified RAD strategy using type IIB restriction endonucleases for genome genotyping, is a relatively simple library preparation scheme (Wang et al., 2016). Compared with the traditional simplified genome technology, the 2b-RAD-seq technique can obtain a larger number of markers with identical length and higher typing accuracy and is widely applied in high-density genetic mapping construction, such as snakehead (*Channa argus*) (Liu et al., 2020), bighead carp (*Aristichthys nobilis*) (Fu et al., 2016), and sea cucumber (*Apostichopus japonicus*) (Tian et al., 2015). Consequently, we used 2b-RAD-seq to construct the genetic linkage map of red swamp crayfish in this study.

The goal of this article was to construct a high-density SNP genetic map by 2b-RAD-seq, screen QTLs associated with growth and sex, and further identify growth- and sex-related candidate genes in red swamp crayfish. Compared with other crustacean species, red swamp crayfish is more complex with 188 chromosomes (Zhang et al., 2018), which indicates that more SNP markers should be developed for high-density map construction. The result of our study will not only reveal the locations and effects of QTLs for growth and sex fundamentally but also provide a basis for marker-assisted breeding and industry promotion in *P. clarkii* in the future.

MATERIALS AND METHODS

Mapping Family Establishment

Wild red swamp crayfish adults were collected from Hongze Lake, Jiangsu, and Poyang Lake, Jiangxi, and were raised in the National Aquatic Biological Resource Center (NABRC). The full-sib families were established by natural mating between one male and four females in one closed culture tank. The genetic relationship among these parents was evaluated using 10 highly polymorphic microsatellite loci developed based on transcriptome data. After PCR amplification, the products were separated by electrophoresis on agarose gels stained with ethidium bromide. Then, the gels were photographed under UV light for band counting, and the bands were binary coded by 1 or 0 for their presence or absence in each genotype. Finally, the genetic distance was calculated by POPGENE software (unpublished data). When the crayfish began to release eggs, every spawning crayfish was transferred to a separate tank for hatching. One family with a high genetic distance was selected as the mapping family. After 3 months of culture, a total of 130 progenies were selected, and six growth-related traits including body weight (BW), full length (FL), body length (BL), tail length (TL), carapaces width (CW), and chelae width (CHW) were measured to compare the growth rate of crayfish; Pearson's correlation coefficient between these traits were also calculated. The sexes of crayfish were identified based on the difference between male and female pleopods. The tail muscle tissues of the parents and 130 progenies were collected and stored in absolute ethyl alcohol for DNA extraction. Total DNA was extracted by using the Genomic DNA Extraction Kit (Promega, USA), following the manufacturer's protocol. DNA quality and concentration were assessed by 1% agarose gel electrophoresis and a UV spectrophotometer and then stored at −80°C for library construction.

2b-RAD Sequencing and *De Novo* Genotyping

A total of 130 offspring and two parents were used for 2b-RAD-seq library construction by the previously published protocols (Wang et al., 2012). In brief, 200 ng genome DNA of each individual was first digested in 15 μl reaction mixture with 1U BsaXI (New England Biolabs) for 45 min at 37°C. Then, 10 μL of digestion product, 0.2 μM of specific adaptors (5 pairs of adaptors

per 5 samples), 0.5 mM ATP (New England Biolabs), 200U T4 DNA ligase (New England Biolabs), 2 μl 10×T4 ligase buffer (New England Biolabs), and 5.9 μl of nuclease-free water were mixed to 20 μl for ligation at 16°C for 1 h. Next, the ligation products were amplified by PCR, and the barcodes were then introduced by specific primers. The amplified products were purified using the MinElute PCR Purification Kit (QIAGEN). Next, all constructed libraries were pooled with equal amounts of product from each library to make a final library, which was then sequenced by the Illumina Hiseq Xten sequencing platform. Raw reads were initially assigned to each individual according to their specific barcodes. Then, the reads were first trimmed to remove adapter sequences and the terminal 3 bp. Next, the reads with no enzyme restriction sites, containing long homopolymers (>10 bp), ambiguous bases (N), or low-quality sequences (quality <20), or of mitochondrial origin were removed. The remaining high-quality reads with a length of 27 bp were used for *de novo* 2b-RAD-seq genotyping.

High-Density Linkage Map Construction

The identified SNP markers that could be segregated in crayfish parents and be genotyped in at least 80% of the offspring were used for map construction. Chi-square tests were performed to check whether the markers were consistent with the expected separation ratio of 1:1 or 1:2:1. The markers with significant segregation distortion ($p < 0.05$) were removed, and slightly distorted markers were remained for map construction. The consensus linkage map was grouped and constructed by the JoinMap 4.0 program (Ooijen and Van, 2006) with a threshold logarithm of odds (LOD) score of 9.5, and the Linkage Map View package in R was used to visualize the mapping results. The expected genetic map length was calculated using *Ge1* (Fishman et al., 2001) and *Ge2* (Chakravarti et al., 1991), and the average of these two indexes was used as the predicted total genetic map length (*Ge*).

QTL Mapping for Growth- and Sex-Related Traits

A QTL analysis of growth- and sex-related traits was performed using MapQTL6 software (Ooijen et al., 2009). Multiple QTL models (MQMs) were used to detect all the significant associations between traits and marker loci (Jansen and Stam, 1994). The LOD scores were calculated at an interval of 1 cM, and the genome-wide (significance level) and linkage group-wide (suggestive level) LOD thresholds were estimated using a permutation test (1,000 replicates) with a confidence interval of 95%. QTL was determined to be significant if the LOD score was higher than the significance threshold estimated by permutation. The percentage of phenotypic variation explained by each QTL was obtained.

Identification and Validation of Potential Candidate Genes

The marker sequences of each QTL region were used to blast against the genome survey (unpublished data provided by Yan-he Li from Huazhong Agricultural University) and annotated by transcriptome data of crayfish and non-redundant protein sequences (NR) from the NCBI database. qPCR was performed on growth and sex candidate genes to reveal the relationship between traits and expression levels. RNA was taken from the tail muscles of three fast-growing (FG) and three slow-growing (SG) crayfish individuals, as well as from the gonads of three male and three female individuals using the SV Total RNA Isolation Kit (Promega, USA). The quality and concentration of RNA were evaluated by 1% agarose gel electrophoresis and a UV spectrophotometer. First-stand cDNAs were synthesized using the PrimeScript™ RT reagent kit (TaKaRa) according to the instructions, and qPCR was performed on the CFX96TM real-time PCR system (Bio-Rad). The *gapdh* gene was selected as the internal control, and all samples were subjected to three technical replicates using the $2^{-\Delta\Delta CT}$ method. Significance test analysis was performed using SPSS19 software.

TABLE 1 | Pearson's correlation between growth traits in F1 progeny (n = 130).

Traits	BW (gm)	FL (cm)	BL (cm)	TL (cm)	CW (cm)	CHW (cm)
BW	1					
FL	0.935	1				
BL	0.937	0.986	1			
TL	0.893	0.959	0.960	1		
CW	0.941	0.983	0.985	0.953	1	
CHW	−0.040	−0.106	−0.100	−0.095	−0.112	1

RESULTS

Characteristics of the Growth- and Sex-Related Traits

In this study, growth traits were measured and sexes were identified in all 130 offspring of the map family. The mean values of BW, FL, BL, TL, CW, and CHW were 2.60 ± 2.22 g, 4.60 ± 1.16 cm, 3.90 ± 0.98 cm, 1.70 ± 0.41 cm, 0.95 ± 0.26 cm, and 0.30 ± 0.13 cm, respectively. Pearson's correlation analysis showed that all traits were highly significantly correlated with each other except chelae width. It indicated that the highest significant correlation value was observed between body length and carapace width ($r = 0.985$), while the correlation values between chela width and all other traits were less than zero (**Table 1**). The mapping family was revealed to consist of 63 males and 67 females with the sex ratio of 1:1.06. Based on morphometric measurement, it was suggested that the growth rate of the males was greater than that of the females under the same culture conditions (**Table 2**). All of these phenotype data related with growth traits were used as mapping panels for QTL analysis.

2b-RAD Sequencing and SNP Genotyping

A total of 348.6 million reads were used, of which 287.4, 30.4, and 30.8 million were from 130 progenies, the male parent, and the female parent, respectively. All of these data have been submitted

TABLE 2 | Comparison of male and female growth traits.

Sex	Number	BW (g)	FL (cm)	BL (cm)	TL (cm)	CW (cm)	CHW (cm)
Female	67	2.23 ± 1.73	4.46 ± 0.64	3.74 ± 0.49	1.64 ± 0.40	0.91 ± 0.25	0.28 ± 0.11
Male	63	2.97 ± 2.62	4.81 ± 1.20	4.03 ± 1.02	1.74 ± 0.44	0.99 ± 0.28	0.32 ± 0.15

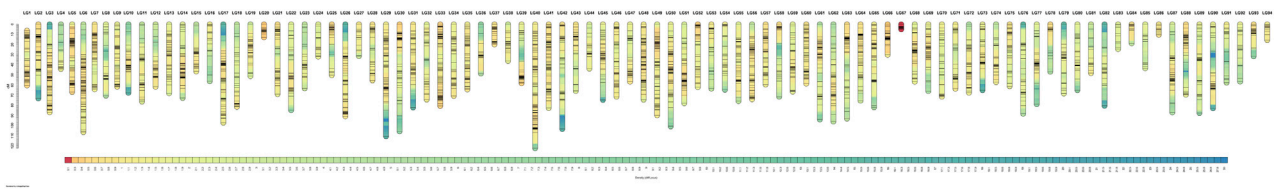

FIGURE 1 | Genetic lengths and marker distributions of 94 linkage groups in the linkage map of red swamp crayfish. The scaleplate on the left indicates genetic distance (centimorgan as the unit). The color module represents the SNP marker distribution density of each centimorgan.

to the Sequence Read Archive (SRA) database of the NCBI (NCBI accession number: PRJNA778942). After sequential quality filtering and sequence trimming, two parents' reads were clustered into 561,392 representative high-quality reference tags. Comparing with the reference sequence, a total of 29,364 SNPs were obtained. These SNPs were divided into three categories: paternal heterozygous (lm×ll, 6,391 SNPs), maternal heterozygous (nn×np, 2,789 SNPs), and both heterozygous (hk×hk, 5,767 SNPs). After filtering the SNPs to remove the segregation distorted markers, the remaining 6,948 loci were used to construct a consensus map.

High-Density Linkage Map

The genetic linkage map, containing 4,878 SNP markers distributed at 4,327 different loci, were successfully divided into 94 LGs, which was consistent with the haploid chromosome number of red swamp crayfish (2n = 188) (Zhang et al., 2018). The consensus map spanned 6,157.737 cM, with an average SNP interval of 1.26 cM. The genetic length of LGs ranged from 5.01 (LG67) to 122.15 (LG40) cM, with an average length of 65.51 cM. The number of markers in each group ranged from 12 (LG83, LG86) to 126 (LG40) with an average of 51.89. The expected map lengths were 6,276.18 cM (*Ge1*) and 6,428.84 cM (*Ge2*), respectively, the average expected map length was 6,352.51 cM (*Ge*), and the genome coverage of this genetic map was 96.93% (**Supplementary Table S1**; **Figure 1**).

QTL Mapping for Growth Traits

QTL analysis of six growth-related traits including BW, FL, BL, TL, CW, and CHW of crayfish was performed using MapQTL6 software. QTL fine mapping based on the abovementioned consensus map showed that a total of 28 group-wide significant QTLs associated with growth traits were identified in nine LGs (LG8, 12, 16, 45, 48, 62, 67, 68, and 70), with LOD scores ranging from 2.32 to 4.38, and a phenotypic variance explained (PVE) ranging from 7.9 to 14.4% (**Table 3**; **Figure 2**). Each QTL interval harbors 1–2 SNP markers. In detail, two QTLs associated with BW including five SNPs were identified on two LGs (LG12 and LG70), with LOD scores ranging from 3.26 to 3.56 and contributed values of phenotypic variance explained (PVE) ranging from 10.9 to 11.8%. For FL, six QTLs including 14 SNPs were identified on six LGs (LG8, LG12, LG16, LG45, LG48, and LG70), and the most significant QTL located at LG12 with the highest LOD value of 4.36 explained 14.4% of PVE. A total of seven QTLs associated with BL were detected on seven LGs (LG8, LG12, LG16, LG45, LG48, LG62, and LG68) containing 19 SNPs with LOD scores ranging from 2.91 to 4.21 and PVE ranging from 9.9 to 14%. For TL, 19 SNPs from five LGs (LG8, LG12, LG16, LG48, and LG67) were identified with LOD scores ranging from 2.32 to 4.38 and PVE ranging from 7.9 to 14.4%. For CW, a total of 19 SNPs from seven LGs (LG8, LG12, LG16, LG45, LG48, LG62, and LG70) were identified, and the highest LOD score was located in LG12 with highest 14% PVE. Four SNPs from two CHW-related QTLs were identified on LG12, and the most important QTL region was located at 11.98–39.603 cM (LOD = 4.17) explaining 14% of the PVE. All the QTLs also indicated that many overlapping regions were identified (**Figure 3**). Notably, LG12 was associated with all growth traits, which suggested that these traits may be regulated by genes on LG12, and LG12 was strongly associated with growth. In addition, a total of 11 genes were detected by searching against the genome of red swamp crayfish using extended SNP markers sequences, such as anti-lipopolysaccharide factor 4 (*alf4*), laminin receptor (*lamr*), golgin subfamily B member 1-like (*golgb1*), TBC1 domain family member 1-like (*tbcd1l*), glutamate receptor subunit 1-like (*glurl*), molt-inhibiting hormone (*mih*), and hyperglycemic peptide 2 precursor (*cprp2*) (**Table 3**).

QTL Mapping for Sex and Screening of Candidate Sex Dimorphic Genes

QTL fine mapping revealed one genome-wide QTL interval in LG20 associated with sex containing a total of 31 SNPs, with LOD scores ranging from 19.71 to 28.60 and PVE ranging from 50.3 to 63.7%. The sex-related QTL interval ranged from 0 to 12.9 cM,

TABLE 3 | Summary information of growth-related QTL and candidate genes.

Trait	LG	QTL	Position (cM)	No. of SNPs	Max LOD	LOD threshold		Max PVE%	Candidate gene name
						Group wide	Genome wide		
BW	12	qBW12	11.980–12.037	2	3.47	3.2	6.9	11.6	Anti-lipopolysaccharide factor 4 (*alf4*)
			17.604	1	3.35	3.2	6.9	11.2	\
	70	qBW70	16.003–18.998	2	3.56	3	6.9	11.8	\
FL	8	qFL8	20.78	1	2.88	2.7	5.3	9.8	\
	12	qFL12	11.980–12.037	2	4.36	3	5.3	14.4	Anti-lipopolysaccharide factor 4 (*alf4*)
			17.604	1	4.06	3	5.3	13.5	\
			31.85	1	3.25	3	5.3	11	\
			37.115–39.603	2	3.33	3	5.3	11.2	\
	16	qFL16	7.995	1	3.48	2.6	5.3	11.7	Laminin receptor (*lamr*)
	45	qFL45	0	1	3.25	3.2	5.3	11	\
	48	qFL48	47.83	1	3.47	3.2	5.3	11.7	Golgin subfamily B member 1-like (*golgb1*)
			50.706	1	3.91	3.2	5.3	13	TBC1 domain family member 1-like (*tbcd1l*)
			54.258	1	3.44	3.2	5.3	11.5	Glutamate (NMDA) receptor subunit 1-like (*glurl*)
			57.525	1	3.8	3.2	5.3	12.7	Molt-inhibiting hormone (*mih*)
	70	qFL70	18.998	1	2.9	2.9	5.3	9.8	\
BL	8	qBL8	20.78	1	3.09	2.6	5.3	10.4	\
	12	qBL12	11.980–12.037	2	3.94	3	5.3	13.1	Anti-lipopolysaccharide factor 4 (*alf4*)
			17.604	1	3.67	3	5.3	12.3	\
			37.115	1	3.06	3	5.3	10.3	\
	16	qBL16	7.995	1	3.39	2.6	5.3	11.4	Laminin receptor (*lamr*)
	45	qBL45	0	1	3.5	3	5.3	11.8	\
	48	qBL48	43.314–44.424	2	3.53	3.1	5.3	11.8	Hyperglycemic peptide 2 precursor (cprp2)
			44.601–44.902	2	3.5	3.1	5.3	11.8	Xpa xeroderma pigmentosum, complementation group A (*xpa*)
			47.83	1	3.94	3.1	5.3	13.1	Golgin subfamily B member 1-like (*golgb1*)
			50.706	1	4.21	3.1	5.3	14	TBC1 domain family member 1-like (*tbcd1l*)
			54.258	1	3.69	3.1	5.3	12.4	Glutamate (NMDA) receptor subunit 1-like (*glurl*)
			57.525	1	3.64	3.1	5.3	12.2	Molt-inhibiting hormone (*mih*)
	62	qBL62	39.552	1	3.07	3	5.3	10.4	Nucleosome-remodeling factor subunit NURF301-like (*nurf301*)
			39.934–40.193	2	3.03	3	5.3	10.3	RNA-directed DNA polymerase from mobile element jockey-like
	68	qBL68	38.682	1	2.91	2.9	5.3	9.9	\
TL	8	qTL8	20.78	1	2.82	2.7	5.3	9.5	\
	12	qTL12	11.980–12.037	2	3.59	2.9	5.3	11.9	Anti-lipopolysaccharide factor 4 (*alf4*)
			17.604	1	3.26	2.9	5.3	10.9	\
			31.85	1	2.99	2.9	5.3	10.1	\
			37.115–39.603	2	3.04	2.9	5.3	10.2	\
	16	qTL16	7.995	1	2.99	2.4	5.3	10.1	Laminin receptor (*lamr*)
	48	qTL48	43.314–44.424	2	3.51	3.1	5.3	11.7	Hyperglycemic peptide 2 precursor (cprp2)
			44.601–44.902	2	3.5	3.1	5.3	11.7	Xpa xeroderma pigmentosum, complementation group A (*xpa*)
			47.83	1	3.87	3.1	5.3	12.8	Golgin subfamily B member 1-like (*golgb1*)
			50.706	1	4.38	3.1	5.3	14.4	TBC1 domain family member 1-like (*tbcd1l*)
			54.258	1	3.78	3.1	5.3	12.5	Glutamate (NMDA) receptor subunit 1-like (*glurl*)
			56.585	1	3.39	3.1	5.3	11.3	Enoyl-CoA hydratase (*ech*)
			57.525–57.806	2	4.12	3.1	5.3	13.6	Molt-inhibiting hormone (*mih*)
	67	qTL67	3.473	1	2.32	2.3	5.3	7.9	\
CW	8	qCW8	20.78	1	3.17	2.7	5.3	10.6	\
	12	qCW12	11.980–12.037	2	4.19	3.1	5.3	13.8	Anti-lipopolysaccharide factor 4 (*alf4*)
			17.604	1	3.94	3.1	5.3	13	\
			31.85	1	3.2	3.1	5.3	10.7	\
			37.115–39.603	2	3.42	3.1	5.3	11.4	\
	16	qCW16	7.995	1	3.04	2.6	5.3	10.2	Laminin receptor (*lamr*)
	45	qCW45	0	1	3.31	3	5.3	11	\
	48	qCW48	47.83	1	3.55	3.4	5.3	11.8	Golgin subfamily B member 1-like (*golgb1*)
			50.706	1	3.93	3.4	5.3	13	TBC1 domain family member 1-like (*tbcd1l*)
			54.258	1	3.74	3.4	5.3	12.4	Glutamate (NMDA) receptor subunit 1-like (*glurl*)
			57.525	1	4.26	3.4	5.3	14	Molt-inhibiting hormone (*mih*)

(Continued on following page)

TABLE 3 | (*Continued*) Summary information of growth-related QTL and candidate genes.

Trait	LG	QTL	Position (cM)	No. of SNPs	Max LOD	LOD threshold		Max PVE%	Candidate gene name
						Group wide	Genome wide		
	62	qCW62	39.552	1	3.32	2.9	5.3	11.1	Nucleosome-remodeling factor subunit NURF301-like (*nurf301*)
			39.934–40.193	2	3.25	2.9	5.3	10.9	RNA-directed DNA polymerase from mobile element jockey-like
	70	qCW70	18.998	1	3.06	2.9	5.3	10.3	\
CHW	12	qCHW12	11.980–12.037	2	4.17	3	6	14	Anti-lipopolysaccharide factor 4 (*alf4*)
			17.604	2	3.91	3	6	13.2	
			39.603	1	3	3	6	10.3	

and the most significant QTL region was located on LG20 at 12.841–12.9 cM with the highest LOD score of 28.6, explaining 63.7% of the PVE (**Figure 4**). In particular, 28 of 31 SNP markers were female heterozygotes (genotype nn×np), which suggested that crayfish might have a female heterogametic ZW sex determination system.

A total of 13 potential candidate genes were further identified from sex-related QTL intervals by searching against the genome of red swamp crayfish using extended SNP markers sequences (**Table 4**; **Figure 5**), such as vacuolar protein sorting-associated protein 4B-like (*vps4bl*), acyl-coenzyme A thioesterase 1 (*acot1*), anti-lipopolysaccharide factor 4 (*alf4*), molt-inhibiting hormone (*mih*), tetratricopeptide repeat protein 21B-like (*ttc21bl*), and tigger transposable element-derived protein 1-like (*tigd1*). It is worth noting that *mih* and *alf4* are also growth-related genes from QTL mapping for growth traits, suggesting a possible correlation between growth and gender.

Validation of Candidate Genes by qRT- PCR Analysis

Based on the abovementioned QTL mapping, a total of 11 growth-related and 13 sex-related potential genes were obtained. The obtained genes were subjected to qPCR and showed that six growth-related genes were significantly differentially expressed in crayfish of different sizes, and eight sex-related genes were differentially expressed in male and female crayfish (**Figure 6**).

DISCUSSION

A high-density linkage map is an important and effective tool for QTL fine mapping, comparative genomic analysis, and genome assembly. Compared with former common molecular markers including SSR and AFLP, SNP markers are the most abundant type of genetic markers in the genome with high genetic stability, which are an ideal choice for constructing genetic linkage maps (Berthier-Schaad et al., 2007; Bourgeois et al., 2013). However, it is a great challenge to develop a sufficient number of SNP markers in a relatively large mapping population. 2b-RAD-seq, one of next-generation sequencing (NGS) technologies, is based on type IIB restriction endonuclease enzymatic cleavage of the genome and substantially reduces the complexity of the target genome, especially for the non-model organisms without a reference genome (Wang et al., 2012). To date, high-density linkage maps constructed based on 2b-Rad have been applied in many aquatic animals, such as red-tail catfish (Zhou et al., 2021) and snakehead (Liu et al., 2020). In this study, we constructed the first genetic linkage map of red swamp crayfish by the 2b-RAD-seq technique. The genetic linkage map contained 4,878 SNP markers with an overall genetic length of 6,157.737 cM and an average marker interval of 1.26 cM, as well as 96.93% genome coverage, which could be considered as a high-density genetic map. To date, this is the first report on genetic linkage map of red swamp crayfish. Furthermore, all the SNP markers were assigned to 94 linkage groups, which indicated that the constructed genetic linkage map of red swamp crayfish contained the largest number of linkage groups compared with the previously reported genetic map.

Growth is one of the important economic traits and priority traits for genetic improvement (Bjarne, 1986; Gui and Zhu, 2012). Traditional genetic improvement strategies for growth traits mainly rely on phenotypic and individual selection, which has a long cycle time and low selection efficiency (Tong and Sun, 2015). Meanwhile, molecular marker-assisted breeding uses the characteristics of molecular markers closely linked to the genes determining the target traits to achieve the purpose of selecting the target traits by detecting molecular markers that can save time and cost and improve the selection efficiency (Collard and Mackill, 2008). QTL mapping based on a high-density linkage map could identify potential candidate genes related with the growth traits in many fish species, which provides an effective approach for breeding programs (Tong and Sun, 2015). In this study, we identified 28 group-wide significant QTLs associated with six growth traits in nine LGs. We also found that most of the QTL loci were overlapped and mainly located in several important linkage groups, which may result from the high phenotypic correlation between these traits of crayfish. It also suggested that these traits may be controlled by some common genes. The same observations were revealed in other species, such as red-tail catfish (Zhou et al., 2021), *Hypophthalmichthys nobilis* (Fu et al., 2016), and *Pungitius pungitius* L. (Laine et al., 2013). In addition, each of the six growth traits had

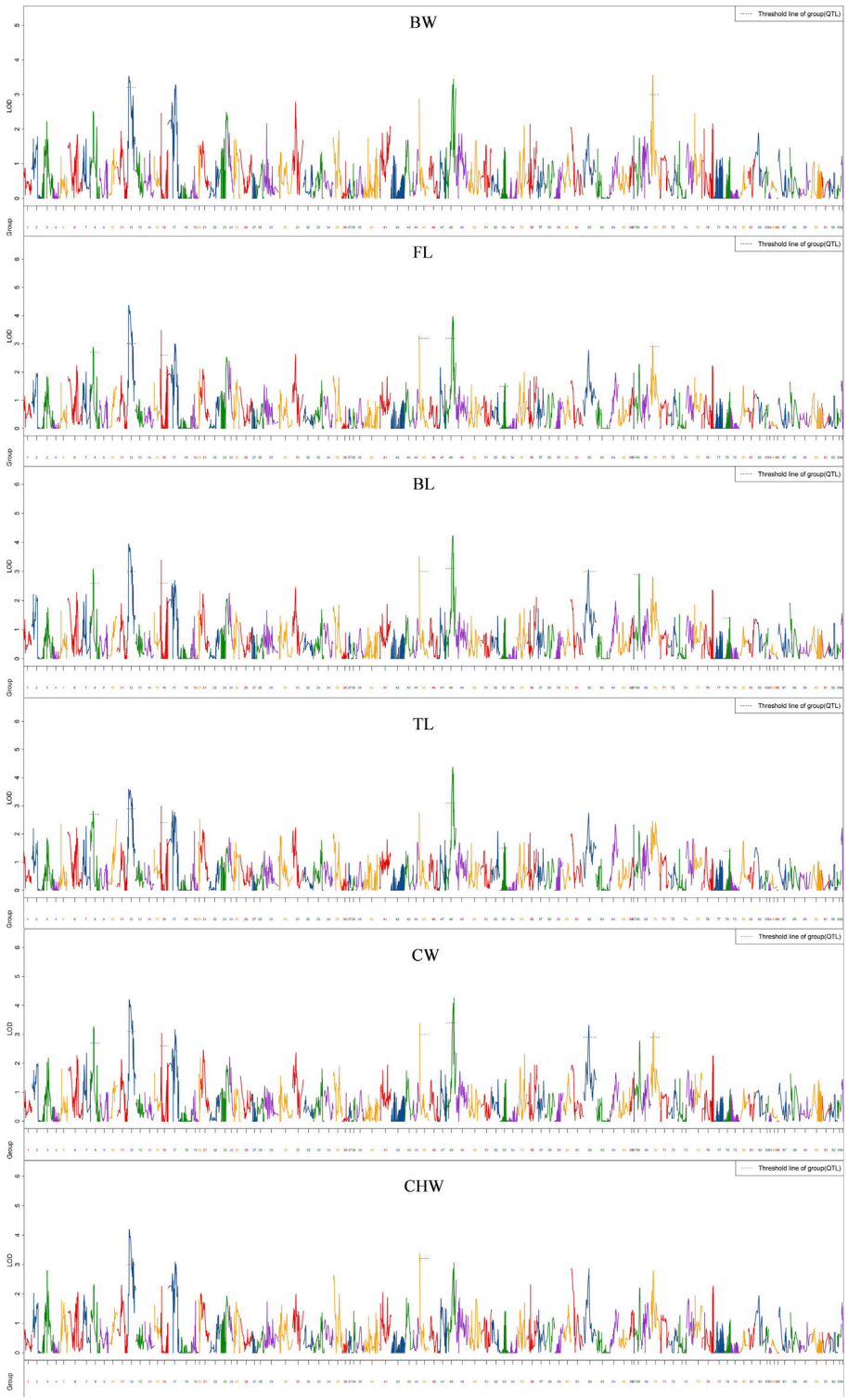

FIGURE 2 | QTL mapping of six growth-related traits of red swamp crayfish. The length of each linkage group is plotted as the coordinate distance in the *x*-axis.

2–7 QTLs with low PVE values of 7.9–14%, suggesting that the growth phenotypic variations of crayfish may be regulated by many QTLs and multiple genes.

A total of 11 candidate growth-related genes from 28 growth-related QTL regions were identified d by QTL mapping. It has been reported that TBC1 domain family member 1 (*tbcd1*) is involved in

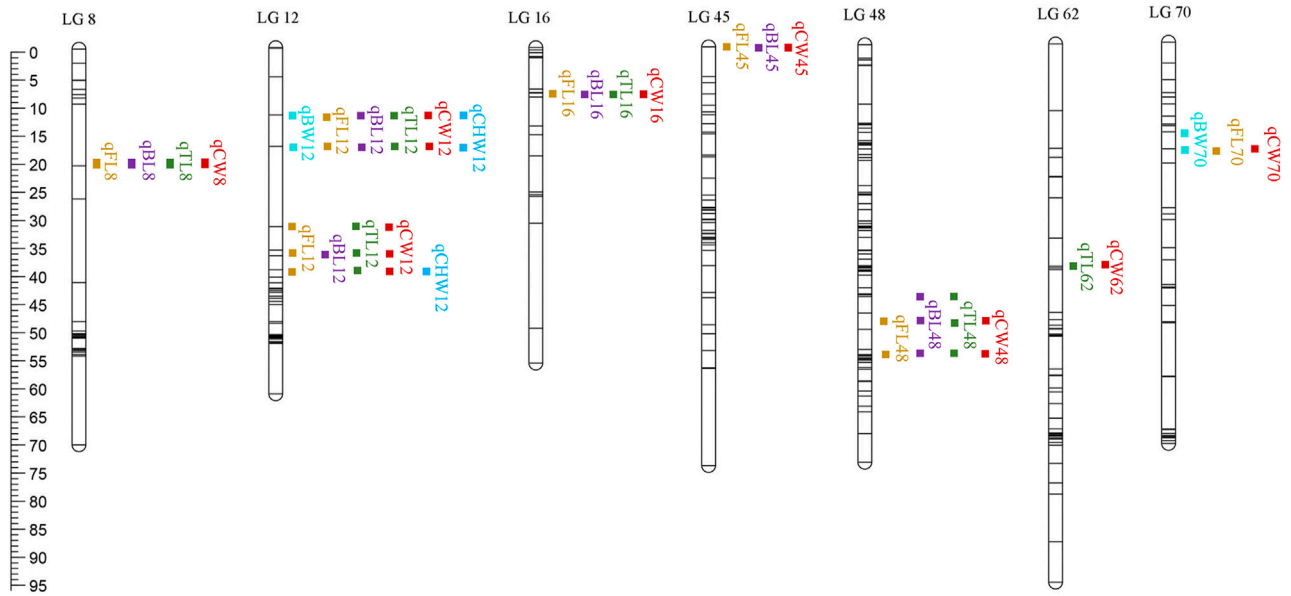

FIGURE 3 | QTL distribution on nine different LGs. The scaleplate on the left indicates genetic distance (centimorgan as the unit).

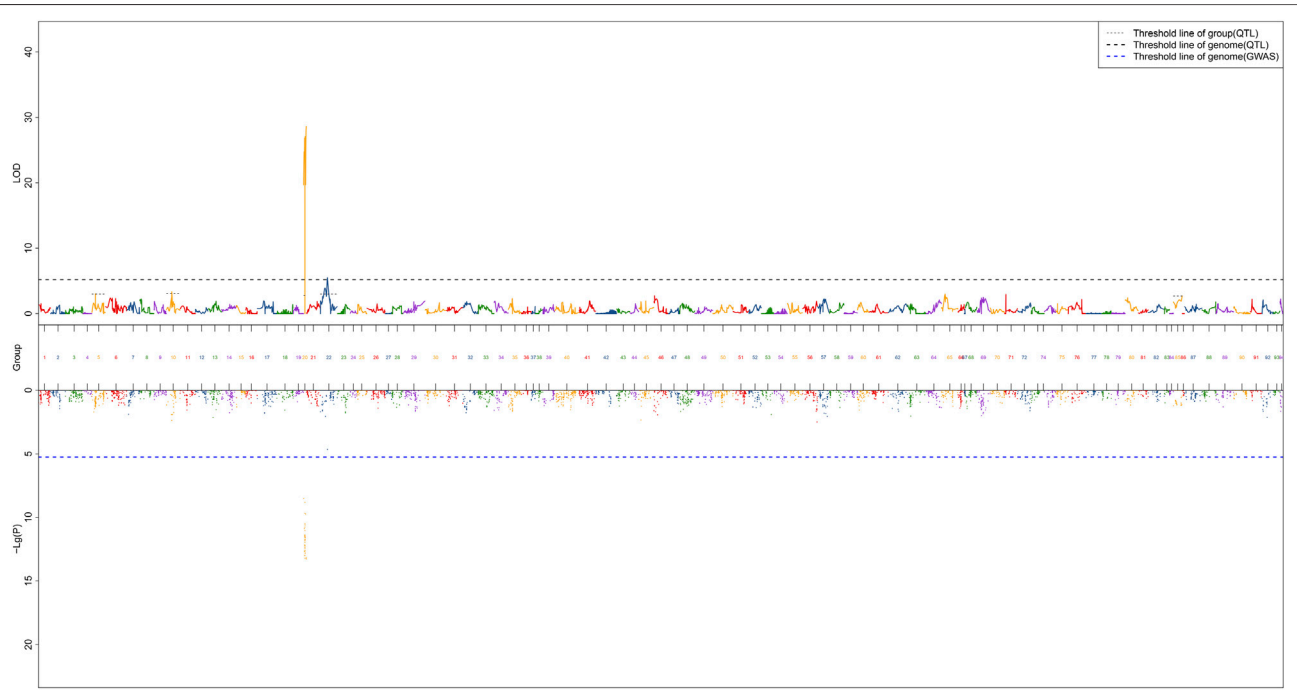

FIGURE 4 | QTL mapping of sex traits of red swamp crayfish. The length of each linkage group is plotted as the coordinate distance in the *x*-axis.

the regulation of glucose transporter protein 4 transport and glucose uptake in adipocytes and skeletal muscle (Treebak et al., 2010), and downregulation of *tbcd1* is associated with enhanced glucose metabolism in mouse skeletal muscle (Okazaki et al., 2020). Golgin subfamily B member 1 (*golgb1*) encodes golgi-associated large transmembrane protein, and it suggests that a novel 65bp was significantly associated with chicken body weight and significantly associated with neck weight and abdominal fat weight (Linstedt and Hauri, 1993; Fu et al., 2020). Enoyl-CoA hydratase (*ech*) catalyzes the second step of an important physiological oxidative pathway in fatty

TABLE 4 | Summary information of sex-related QTL and candidate genes.

LG	QTL	Position (cM)	No. of SNPs	Max LOD	LOD threshold		Max PVE%	Candidate gene name
					Group wide	Genome wide		
20	qSex20	0–3.571	2	24.02	2.8	5.2	57.3	Vacuolar protein sorting-associated protein 4B-like (*vps4bl*)
		3.855–5.609	4	26.81	2.8	5.2	61.3	Small GTPase Rab5 (*rab5*)
		5.827–6.328	3	26.55	2.8	5.2	61	Anti-lipopolysaccharide factor 4 (*alf4*)
		6.344–6.350	3	26.55	2.8	5.2	61	Molt-inhibiting hormone (*mih*)
		6.353–6.398	4	26.55	2.8	5.2	61	Acyl-coenzyme A thioesterase 1 (*acot1*)
		6.402	1	25.4	2.8	5.2	59.3	Prophenoloxidase (*ppo*)
		6.403–6.429	4	25.4	2.8	5.2	59.3	Tetratricopeptide repeat protein 21B-like (*ttc21bl*)
		6.449	1	25.4	2.8	5.2	59.3	Anti-lipopolysaccharide factor 4 (*alf4*)
		6.49–7.041	2	25.4	2.8	5.2	59.3	Zinc finger MYM-type protein 1-like (*zmym1l*)
		7.227	1	27.05	2.8	5.2	61.6	Tigger transposable element-derived protein 1-like (*tigd1*)
		7.24	1	27.05	2.8	5.2	61.6	PAB-dependent poly(A)-specific ribonuclease subunit PAN2 (*pan2*)
		7.44	1	27.05	2.8	5.2	61.6	Ecdysone receptor 2 (*ecr2*)
		7.983	1	25.4	2.8	5.2	59.3	Hyperglycemic peptide 2 precursor (*cprp2*)
		9.679	1	19.71	2.8	5.2	50.3	Spermatogonial stem-cell renewal factor (*ssrf*)
		12.841–12.9	2	28.6	2.8	5.2	63.7	\

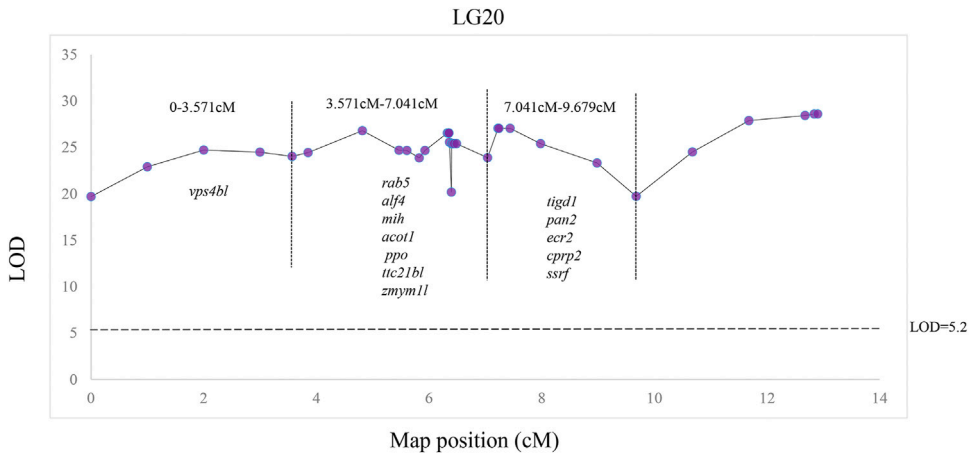

FIGURE 5 | Distribution of 2b-RAD-seq SNP markers and candidate sex dimorphic genes in LG20.

acid metabolism, and it is involved in growth and development (Agnihotri and Liu, 2003). Crustacean growth is discontinuous and greatly influenced by the molting cycle, and it has been revealed that *mih* mainly regulates the molting process by inhibiting the synthesis of ecdysteroid (Jung et al., 2013). Moreover, SNPs in *mih* were found to be significantly associated with growth traits in both white shrimp and red swamp crayfish (Li et al., 2011; Xu et al., 2019). Moreover, *nurf301* is required for ecdysteroid signaling and metamorphosis, and it is suggested that it is a direct effector of nuclear receptor activity (Badenhorst et al., 2005; Nakatsuji et al., 2009). In addition, qPCR results showed that *nurf301*, *glur*, *lamr*, *pol*, *xpa*, and *alf4* were expressed differently in FG and SG groups. Therefore, we hypothesized that these genes also regulate growth and development in crayfish.

Many aquatic animals display significant growth differences between males and females; therefore, mono-sex population culture is an effective way to improve economic value (Mei and Gui, 2015; Gui et al., 2022). Sex determination is an important research field in developmental biology and evolutionary biology, and it is a plastic developmental process (Charlesworth and Mank, 2010; Li and Gui, 2018). Sex determination in crustaceans shows degrees of plasticity, being influenced both genetic and other epigenetic factors including environmental variables such as light and temperature, parasites, and even diet (Ford, 2008). Although the sex determination system of some crustaceans was revealed by karyotype analysis (Lécher et al., 1995), it is too difficult to distinguish sex chromosomes by karyotype analysis of the sex determination mechanism due to super genome complexity such as small size, large number, and highly condensed chromosomes in red swamp crayfish (Torrecilla et al., 2017). Therefore, sex-related QTL mapping and sex-specific marker identification are potential

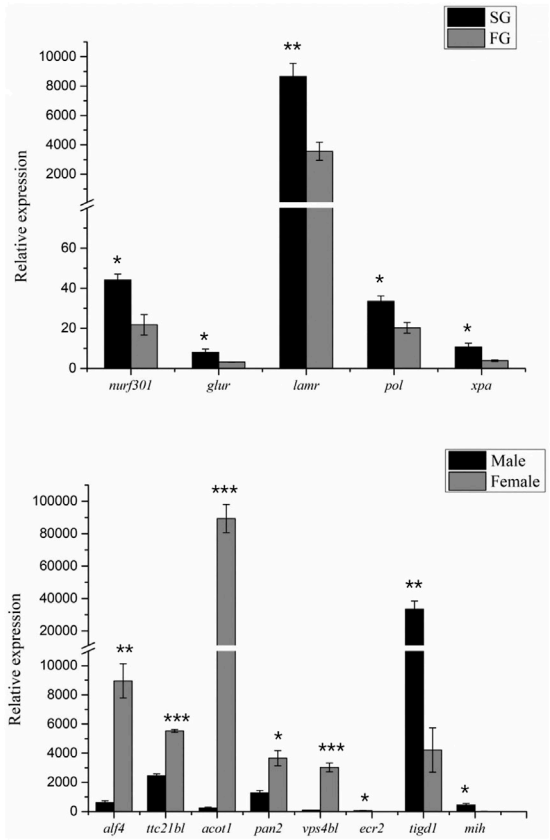

FIGURE 6 | qPCR validation of growth and sex candidate genes. "***," significant at $p < 0.001$. "**," significant at $p < 0.01$. "*," significant at $p < 0.05$.

methods for sex-associated gene identification and sex determination mechanism uncovering. Sex determination systems of some crustaceans have been revealed by QTL mapping, such as the XX/XY sex determination system in swimming crab (*Portunus trituberculatus*) (Lv et al., 2018) and the ZZ/ZW sex determination system in black tiger shrimp (*Penaeus monodon*) (Guo et al., 2019). In this study, one genome-wide significant QTL related to sex was successfully detected in LG20, explaining 50.3–63.7% of the phenotypic variation, which suggested that sex determination is a single-point trait and employs a specific sex chromosome in red swamp crayfish. In particular, 28 of 31 identified markers were female heterozygotes (genotype nn×np), strongly suggesting that red swamp crayfish may have the female heterogametic ZZ/ZW sex determination system.

Sex determination genes are conducive to uncover the sex characteristics in aquatic animals; therefore, some sex determination genes have been identified (Bao et al., 2019; Trant et al., 2001; Matsuda et al., 2002; Yokoi et al., 2002; Zhou et al., 2002). However, no sex determination genes have been identified in red swamp crayfish to date. In this study, a total of 13 potential genes were detected within the sex QTL region, for example, *ppo*, *vps4bl, acot1, rab5, ttc21bl, and ssrf* (**Table 4**). Prophenoloxidase encoded by *ppo* was detected to be different between the females and the males in *Callosobruchus maculatus*is, but it was suggested that prophenoloxidase was an important defense for invertebrates. We also identified a spermatogonial stem-cell renewal factor (*ssrf*), which could induce spermatogonial mitosis to maintain long-term fertility and sustain sufficiently high levels of spermatogenesis (Mäkelä and Hobbs, 2019). Interestingly, *mih* and *alf4* were identified in both growth-related and sex-related QTLs, which suggested an inevitable correlation existed between growth and sex. Although these sexually dimorphic genes expressed differentially between the males and the females by qPCR validation (**Figure 6**), they have not been revealed to be necessary for sex determination in other aquatic animals. So, we predicted that the sex determination gene in red swamp crayfish might be other novel genes and it needed more comprehensive and systematic research in the future.

CONCLUSION

In this study, 4878 SNP markers distributed in 94 LGs were developed to construct a consensus linkage map, which represents the largest linkage map to date. Abundant QTLs related to growth and sex were identified, uncovering the sexual growth dimorphism and sex determination mechanism. Moreover, our findings will provide effective markers for MAS in red swamp crayfish 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/, PRJNA778942.

ETHICS STATEMENT

The animal study was reviewed and approved by the Laboratory Animals of the Institute of Hydrobiology, the Chinese Academy of Sciences, China.

AUTHOR CONTRIBUTIONS

ZW and JG contributed to the study design. XG performed the experiments and drafted the manuscript. YZ, ML, and ZL helped in the execution of some experiments. ZW, LZ, and JG revised the manuscript. All authors read the manuscript, gave suggestions and comments for the improvement, and approved the final manuscript.

FUNDING

This work was supported by grants from the National Key R and D Program of China (2018YFD0901201) and Key R&D projects in Hubei Province (2021BBA232).

ACKNOWLEDGMENTS

The authors would like to thank Zhong-hu Tao and Na-na Shu from the Qianjiang Fisheries Technology Extension Center for sample collection.

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.852280/full#supplementary-material

REFERENCES

Agnihotri, G., and Liu, H.-w. (2003). Enoyl-CoA Hydratase. Reaction, Mechanism, and Inhibition. *Bioorg. Med. Chem.* 11, 9–20. doi:10.1016/s0968-0896(02)00333-4

Badenhorst, P., Xiao, H., Cherbas, L., Kwon, S. Y., Voas, M., Rebay, I., et al. (2005). The Drosophila Nucleosome Remodeling Factor NURF Is Required for Ecdysteroid Signaling and Metamorphosis. *Genes Dev.* 19, 2540–2545. doi:10.1101/gad.1342605

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

Bao, L., Tian, C., Liu, S., Zhang, Y., Elaswad, A., Yuan, Z., et al. (2019). The Y Chromosome Sequence of the Channel Catfish Suggests Novel Sex Determination Mechanisms in Teleost Fish. *BMC Biol.* 17, 6. doi:10.1186/s12915-019-0627-7

Berthier-Schaad, Y., Kao, W. H. L., Coresh, J., Zhang, L., Ingersoll, R. G., Stephens, R., et al. (2007). Reliability of High-Throughput Genotyping of Whole Genome Amplified DNA in SNP Genotyping Studies. *Electrophoresis* 28, 2812–2817. doi:10.1002/elps.200600674

Bjarne, G. (1986). Growth and Reproduction in Fish and Shellfish. *Aquaculture* 57, 37–55. doi:10.1016/0044-8486(86)90179-1

Bourgeois, Y. X. C., Lhuillier, E., Cézard, T., Bertrand, J. A. M., Delahaie, B., Cornuault, J., et al. (2013). Mass Production of SNP Markers in a Nonmodel Passerine Bird through RAD Sequencing and Contig Mapping to the Zebra Finch Genome. *Mol. Ecol. Resour.* 13, 899–907. doi:10.1111/1755-0998.12137

Celi, M., Filiciotto, F., Parrinello, D., Buscaino, G., Damiano, A., Cuttitta, A., et al. (2013). Physiological and Agonistic Behavioural Response of *Procambarus clarkii* to an Acoustic Stimulus. *J. Exp. Biol.* 216, 709–718. doi:10.1242/jeb.078865

Chakravarti, A., Lasher, L. K., and Reefer, J. E. (1991). A Maximum Likelihood Method for Estimating Genome Length Using Genetic Linkage Data. *Genetics* 128, 175–182. doi:10.1093/genetics/128.1.175

Charlesworth, D., and Mank, J. E. (2010). The Birds and the Bees and the Flowers and the Trees: Lessons from Genetic Mapping of Sex Determination in Plants and Animals. *Genetics* 186, 9–31. doi:10.1534/genetics.110.117697

Chen, A.-J., Gao, L., Wang, X.-W., Zhao, X.-F., and Wang, J.-X. (2013). SUMO-conjugating Enzyme E2 UBC9 Mediates Viral Immediate-Early Protein SUMOylation in Crayfish to Facilitate Reproduction of white Spot Syndrome Virus. *J. Virol.* 87, 636–647. doi:10.1128/JVI.01671-12

Collard, B. C. Y., and Mackill, D. J. (2008). Marker-assisted Selection: an Approach for Precision Plant Breeding in the Twenty-First century. *Phil. Trans. R. Soc. B* 363, 557–572. doi:10.1098/rstb.2007.2170

Davey, J. W., and Blaxter, M. L. (2010). RADSeq: Next-Generation Population Genetics. *Brief. Funct. Genomics* 9, 416–423. doi:10.1093/bfgp/elq031

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

Feng, X., Yu, X., Fu, B., Wang, X., Liu, H., Pang, M., et al. (2018). A High-Resolution Genetic Linkage Map and QTL fine Mapping for Growth-Related Traits and Sex in the Yangtze River Common Carp (*Cyprinus carpio Haematopterus*). *BMC Genomics* 19, 230. doi:10.1186/s12864-018-4613-1

Fishman, L., Kelly, A. J., Morgan, E., and Willis, J. H. (2001). A Genetic Map in the *Mimulus Guttatus* Species Complex Reveals Transmission Ratio Distortion Due to Heterospecific Interactions. *Genetics* 159, 1701–1716. doi:10.1093/genetics/159.4.1701

Ford, A. T. (2008). Can You Feminise a Crustacean? *Aquat. Toxicol.* 88, 316–321. doi:10.1016/j.aquatox.2008.04.013

Fu, B., Liu, H., Yu, X., and Tong, J. (2016). A High-Density Genetic Map and Growth Related QTL Mapping in Bighead Carp (*Hypophthalmichthys Nobilis*). *Sci. Rep.* 6, 28679. doi:10.1038/srep28679

Fu, R., Ren, T., Li, W., Liang, J., Mo, G., Luo, W., et al. (2020). A Novel 65-bp Indel in the *GOLGB1* Gene Is Associated with Chicken Growth and Carcass Traits. *Animals* 10, 475. doi:10.3390/ani10030475

Gui, J., and Zhu, Z. (2012). Molecular Basis and Genetic Improvement of Economically Important Traits in Aquaculture Animals. *Chin. Sci. Bull.* 57, 1751–1760. doi:10.1007/s11434-012-5213-0

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. Security* 1, 100002–100019. doi:10.1016/j.watbs.2021.11.001

Guo, L., Xu, Y.-H., Zhang, N., Zhou, F.-L., Huang, J.-H., Liu, B.-S., et al. (2019). A High-Density Genetic Linkage Map and QTL Mapping for Sex in Black Tiger Shrimp (*Penaeus monodon*). *Front. Genet.* 10, 326. doi:10.3389/fgene.2019.00326

Hobbs, H. H., Jass, J. P., and Huner, J. V. (1989). A Review of Global Crayfish Introductions with Particular Emphasis on Two north American Species *(Decapoda, Cambaridae)*. *Crustac* 56, 299–316. doi:10.1163/156854089X00275

Jansen, R. C., and Stam, P. (1994). High Resolution of Quantitative Traits into Multiple Loci via Interval Mapping. *Genetics* 136, 1447–1455. doi:10.1093/genetics/136.4.1447

Jung, H., Lyons, R. E., Hurwood, D. A., and Mather, P. B. (2013). Genes and Growth Performance in Crustacean Species: a Review of Relevant Genomic Studies in Crustaceans and Other Taxa. *Rev. Aquacult.* 5, 77–110. doi:10.1111/raq.12005

Laine, V. N., Shikano, T., Herczeg, G., Vilkki, J., and Merilä, J. (2013). Quantitative Trait Loci for Growth and Body Size in the Nine-Spined sticklebackPungitius pungitiusL. *Mol. Ecol.* 22, 5861–5876. doi:10.1111/mec.12526

Lécher, P., Defaye, D., and Noel, P. (1995). Chromosomes and Nuclear DNA of Crustacea. *Invertebrate Reprod. Dev.* 27, 85–114. doi:10.1080/07924259.1995.9672440

Li, X.-Y., and Gui, J.-F. (2018). Diverse and Variable Sex Determination Mechanisms in Vertebrates. *Sci. China Life Sci.* 61, 1503–1514. doi:10.1007/s11427-018-9415-7

Li, F., Xin, J. J., Feng, T. T., and Liu, X. L. (2011). Evidence for a Common Evolutionary Origin of Coronavirus Spike Protein Receptor-Binding Subunits. *J. Virol.* 86, 2856–2858. doi:10.3923/javaa.2011.2856.285810.1128/jvi.06882-11

Lin, L.-J., Chen, Y.-J., Chang, Y.-S., and Lee, C.-Y. (2013). Neuroendocrine Responses of a Crustacean Host to Viral Infection: Effects of Infection of white Spot Syndrome Virus on the Expression and Release of Crustacean Hyperglycemic Hormone in the Crayfish *Procambarus clarkii*. *Comp. Biochem. Physiol. A: Mol. Integr. Physiol.* 164, 327–332. doi:10.1016/j.cbpa.2012.11.009

Linstedt, A. D., and Hauri, H. P. (1993). Giantin, a Novel Conserved Golgi Membrane Protein Containing a Cytoplasmic Domain of at Least 350 kDa. *MBoC* 4, 679–693. doi:10.1091/mbc.4.7.679

Liu, H., Fu, B., Pang, M., Feng, X., Yu, X., and Tong, J. (2017). A High-Density Genetic Linkage Map and QTL fine Mapping for Body Weight in Crucian Carp (*Carassius auratus*) Using 2b-RAD Sequencing. *G3 (Bethesda)* 7, 2473–2487. doi:10.1534/g3.117.041376

Liu, H., Luo, Q., Ou, M., Zhu, X., Zhao, J., and Chen, K. (2020). High-density Genetic Linkage Map and QTL fine Mapping of Growth and Sex in Snakehead (*Channa Argus*). *Aquaculture* 519, 734760. doi:10.1016/j.aquaculture.2019.734760

Lv, J., Sun, D., Huan, P., Song, L., Liu, P., and Li, J. (2018). QTL Mapping and Marker Identification for Sex-Determining: Indicating XY Sex Determination System in the Swimming Crab (*Portunus Trituberculatus*). *Front. Genet.* 9, 337. doi:10.3389/fgene.2018.00337

Mäkelä, J.-A., and Hobbs, R. M. (2019). Molecular Regulation of Spermatogonial Stem Cell Renewal and Differentiation. *Reproduction* 158, R169–R187. doi:10.1530/REP-18-0476

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., et al. (2002). DMY Is a Y-specific DM-Domain Gene Required for Male Development in the Medaka Fish. *Nature* 417, 559–563. doi:10.1038/nature751

Mei, J., and Gui, J.-F. (2015). Genetic Basis and Biotechnological Manipulation of Sexual Dimorphism and Sex Determination in Fish. *Sci. China Life Sci.* 58, 124–136. doi:10.1007/s11427-014-4797-9

Nakatsuji, T., Lee, C.-Y., and Watson, R. D. (2009). Crustacean Molt-Inhibiting Hormone: Structure, Function, and Cellular Mode of Action. *Comp. Biochem. Physiol. Part A: Mol. Integr. Physiol.* 152, 139–148. doi:10.1016/j.cbpa.2008.10.012

Okazaki, Y., Murray, J., Ehsani, A., Clark, J., Whitson, R. H., Hirose, L., et al. (2020). Increased Glucose Metabolism in Arid5b−/− Skeletal Muscle Is Associated with the Down-Regulation of TBC1 Domain Family Member 1 (TBC1D1). *Biol. Res.* 53, 45. doi:10.1186/s40659-020-00313-3

Ooijen, J., and Van, J. W. (2006). *JoinMap 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyazma BV.

Ooijen, J. V., Ooijen, J. V., Ooijen, J., Hoorn, J., Duin, J., and Van, J. W. (2009). *MapQTL®6. Software for the Mapping of Quantitative Trait Loci in Experimental Populations of Diploid Species*. Wageningen: Kyazma BV.

Peng, W., Xu, J., Zhang, Y., Feng, J., Dong, C., Jiang, L., et al. (2016). An Ultra-high Density Linkage Map and QTL Mapping for Sex and Growth-Related Traits of Common Carp (*Cyprinus carpio*). *Sci. Rep.* 6, 26693. doi:10.1038/srep26693

Raina, V. S., Kour, A., Chakravarty, A. K., and Vohra, V. (2020). Marker-assisted Selection Vis-à-Vis Bull Fertility: Coming Full circle-a Review. *Mol. Biol. Rep.* 47, 9123–9133. doi:10.1007/s11033-020-05919-0

Shen, H., Hu, Y., Ma, Y., Zhou, X., Xu, Z., Shui, Y., et al. (2014). In-depth Transcriptome Analysis of the Red Swamp Crayfish *Procambarus clarkii*. *PLoS One* 9, e110548. doi:10.1371/journal.pone.0110548

Shi, Y., Wang, S., Gu, Z., Lv, J., Zhan, X., Yu, C., et al. (2014). High-density Single Nucleotide Polymorphisms Linkage and Quantitative Trait Locus Mapping of the Pearl Oyster, Pinctada Fucata Martensii Dunker. *Aquaculture* 434, 376–384. doi:10.1016/j.aquaculture.2014.08.044

Tian, M., Li, Y., Jing, J., Mu, C., Du, H., Dou, J., et al. (2015). Construction of a High-Density Genetic Map and Quantitative Trait Locus Mapping in the Sea Cucumber *Apostichopus Japonicus*. *Sci. Rep.* 5, 14852. doi:10.1038/srep14852

Tong, J., and Sun, X. (2015). Genetic and Genomic Analyses for Economically Important Traits and Their Applications in Molecular Breeding of Cultured Fish. *Sci. China Life Sci.* 58, 178–186. doi:10.1007/s11427-015-4804-9

Torrecilla, Z., Martínez-Lage, A., Perina, A., González-Ortegón, E., and González-Tizón, A. M. (2017). Comparative Cytogenetic Analysis of marine *Palaemon* Species Reveals a X1X1X2X2/X1X2Y Sex Chromosome System in *Palaemon Elegans*. *Front. Zool.* 14, 47. doi:10.1186/s12983-017-0233-x

Trant, J. M., Gavasso, S., Ackers, J., Chung, B.-C., and Place, A. R. (2001). Developmental Expression of Cytochrome P450 Aromatase Genes (CYP19a and CYP19b) in Zebrafish Fry (*Danio rerio*). *J. Exp. Zool.* 290, 475–483. doi:10.1002/jez.1090

Treebak, J. T., Taylor, E. B., Witczak, C. A., An, D., Toyoda, T., Koh, H.-J., et al. (2010). Identification of a Novel Phosphorylation Site on TBC1D4 Regulated by AMP-Activated Protein Kinase in Skeletal Muscle. *Am. J. Physiology-Cell Physiol.* 298, C377–C385. doi:10.1152/ajpcell.00297.2009

Wang, S., Meyer, E., McKay, J. K., and Matz, M. V. (2012). 2b-RAD: a Simple and Flexible Method for Genome-wide Genotyping. *Nat. Methods* 9, 808–810. doi:10.1038/nmeth.2023

Wang, S., Liu, P., Lv, J., Li, Y., Cheng, T., Zhang, L., et al. (2016). Serial Sequencing of Isolength RAD Tags for Cost-Efficient Genome-wide Profiling of Genetic and Epigenetic Variations. *Nat. Protoc.* 11, 2189–2200. doi:10.1038/nprot.2016.133

Wu, X., Xiong, H., Wang, Y., and Du, H. (2012). Evidence for Cell Apoptosis Suppressing white Spot Syndrome Virus Replication in *Procambarus clarkii* at High Temperature. *Dis. Aquat. Org.* 102, 13–21. doi:10.3354/dao02532

Xu, Y., Peng, G., Sun, M., Li, J., Yan, W., Tang, J., et al. (2019). Genomic Organization of the Molt-Inhibiting Hormone Gene in the Red Swamp Crayfish *Procambarus clarkii* and Characterization of Single-Nucleotide Polymorphisms Associated with Growth. *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* 237, 110334. doi:10.1016/j.cbpb.2019.110334

Yokoi, H., Kobayashi, T., Tanaka, M., Nagahama, Y., Wakamatsu, Y., Takeda, H., et al. (2002). sox9 in a Teleost Fish, Medaka (*Oryzias latipes*): Evidence for Diversified Function ofSox9 in Gonad Differentiation. *Mol. Reprod. Dev.* 63, 5–16. doi:10.1002/mrd.10169

Yu, Y., Zhang, X., Yuan, J., Li, F., Chen, X., Zhao, Y., et al. (2015). Genome Survey and High-Density Genetic Map Construction Provide Genomic and Genetic Resources for the pacific white Shrimp *Litopenaeus Vannamei*. *Sci. Rep.* 5, 15612. doi:10.1038/srep15612

Yue, G. H., Wang, G. L., Zhu, B. Q., Wang, C. M., Zhu, Z. Y., and Lo, L. C. (2008). Discovery of Four Natural Clones in a Crayfish Species *Procambarus clarkii*. *Int. J. Biol. Sci.* 4, 279–282. doi:10.7150/ijbs.4.279

Yue, G. H. (2014). Recent Advances of Genome Mapping and Marker-Assisted Selection in Aquaculture. *Fish Fish* 15, 376–396. doi:10.1111/faf.12020

Zenger, K. R., Khatkar, M. S., Jerry, D. R., and Raadsma, H. W. (2017). The Next Wave in Selective Breeding: Implementing Genomic Selection in Aquaculture. *Proc. Assoc. Advmt. Anim. Breed. Genet.* 22, 105–112.

Zhang, S., Yu, S, H., and Qiu, G, F. (2018). Chromosome and Karyotype of the Crayfish (*Procambarus clarkii*). *J. Fish. China* 42, 1513–1519. doi:10.11964/jfc.20171211084

Zhang, S., Zhang, X., Chen, X., Xu, T., Wang, M., Qin, Q., et al. (2019). Construction of a High-Density Linkage Map and QTL fine Mapping for Growth- and Sex-Related Traits in Channel Catfish (*Ictalurus punctatus*). *Front. Genet.* 10, 251. doi:10.3389/fgene.2019.00251

Zhou, R., Cheng, H., Zhang, Q., Guo, Y., Cooper, R. K., and Tiersch, T. R. (2002). SRY-related Genes in the Genome of the rice Field Eel (*Monopterus albus*). *Genet. Sel. Evol.* 34, 129–137. doi:10.1186/1297-9686-34-1-129

Zhou, Y.-L., Wang, Z.-W., Guo, X.-F., Wu, J.-J., Lu, W.-J., Zhou, L., et al. (2021). Construction of a High-Density Genetic Linkage Map and fine Mapping of QTLs for Growth and Sex-Related Traits in Red-Tail Catfish (*Hemibagrus Wyckioides*). *Aquaculture* 531, 735892. doi:10.1016/j.aquaculture.2020.735892

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 26 August 2022
DOI 10.3389/fgene.2022.896774

OPEN ACCESS

EDITED BY
Alexandre Wagner Silva Hilsdorf,
University of Mogi das Cruzes, Brazil

REVIEWED BY
Hugo H Montaldo,
National Autonomous University of Mexico, Mexico
Timothy D. Leeds,
United States Department of Agriculture (USDA), United States

*CORRESPONDENCE
Binyam Dagnachew,
Binyam.dagnachew@nofima.no

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

RECEIVED 15 March 2022
ACCEPTED 14 July 2022
PUBLISHED 26 August 2022

CITATION
Dagnachew B, Aslam ML, Hillestad B, Meuwissen T and Sonesson A (2022), Use of DNA pools of a reference population for genomic selection of a binary trait in Atlantic salmon.
Front. Genet. 13:896774.
doi: 10.3389/fgene.2022.896774

Use of DNA pools of a reference population for genomic selection of a binary trait in Atlantic salmon

Binyam Dagnachew[1]*, Muhammad Luqman Aslam[1], Borghild Hillestad[2], Theo Meuwissen[3] and Anna Sonesson[1]

[1]Fisheries and Aquaculture Research, Nofima AS—Norwegian Institute of Food, Tromsø, Norway, [2]Benchmark Genetics, Bergen, Norway, [3]Norwegian University of Life Sciences, As, Norway

Genomic selection has a great potential in aquaculture breeding since many important traits are not directly measured on the candidates themselves. However, its implementation has been hindered by staggering genotyping costs because of many individual genotypes. In this study, we explored the potential of DNA pooling for creating a reference population as a tool for genomic selection of a binary trait. Two datasets from the SalmoBreed population challenged with salmonid alphavirus, which causes pancreas disease, were used. Dataset-1, that includes 855 individuals (478 survivors and 377 dead), was used to develop four DNA pool samples (i.e., 2 pools each for dead and survival). Dataset-2 includes 914 individuals (435 survivors and 479 dead) belonging to 65 full-sibling families and was used to develop in-silico DNA pools. SNP effects from the pool data were calculated based on allele frequencies estimated from the pools and used to calculate genomic breeding values (GEBVs). The correlation between SNP effects estimated based on individual genotypes and pooled data increased from 0.3 to 0.912 when the number of pools increased from 1 to 200. A similar trend was also observed for the correlation between GEBVs, which increased from 0.84 to 0.976, as the number of pools per phenotype increased from 1 to 200. For dataset-1, the accuracy of prediction was 0.71 and 0.70 when the DNA pools were sequenced in 40× and 20×, respectively, compared to an accuracy of 0.73 for the SNP chip genotypes. For dataset-2, the accuracy of prediction increased from 0.574 to 0.691 when the number of in-silico DNA pools increased from 1 to 200. For this dataset, the accuracy of prediction using individual genotypes was 0.712. A limited effect of sequencing depth on the correlation of GEBVs and prediction accuracy was observed. Results showed that a large number of pools are required to achieve as good prediction as individual genotypes; however, alternative effective pooling strategies should be studied to reduce the number of pools without reducing the prediction power. Nevertheless, it is demonstrated that pooling of a reference population can be used as a tool to optimize between cost and accuracy of selection.

KEYWORDS

genomic selection, DNA pooling, reference population, salmon, in-silico

1 Introduction

Genomic selection (GS) is becoming a practical and effective breeding tool for many livestock species because of the rapid development of high-throughput genotyping technologies that reduce genotyping costs. The potential application of GS for aquaculture species has been studied (Nielsen et al., 2009; Sonesson and Meuwissen, 2009), and it showed an increase in genetic gain especially for traits that are difficult to improve by traditional selection such as disease resistance. Therefore, it is of particular interest in aquaculture species as most breeding goal traits in these species are measured on sibs of the selection candidates.

For a conventional GS breeding program, two large datasets are required: a training set (reference population) with genotyped and phenotyped individuals and a prediction set (selection candidates) containing only genotyped individuals (Meuwissen et al., 2001; Goddard and Hayes, 2007). The sizes of these datasets determine the rate of genetic improvement through influencing components of the equation. On one hand, the prediction accuracy relies on the size of the training set to estimate parameters (i.e., marker effects) and the marker density at which reference individuals are genotyped. On the other hand, the size of the prediction set determines the selection intensity and consequently the response to selection. However, increasing either the training set or the prediction set increases the cost of GS programs.

Even though the genotyping cost per individual is reducing, implementation of conventional GS is expensive in aquaculture compared to other livestock species because of the fact that the number of selection candidates and their siblings to genotype is large. Therefore, it is of interest to reduce either the number of individuals or the number of markers to genotype without reducing the prediction accuracy significantly. Strategies for reducing the number of markers have been described in many studies (Lillehammer et al., 2013; Ødegård and Meuwissen, 2014; Dagnachew and Meuwissen, 2019; Tsairidou et al., 2020). This study investigates the impact of reducing the number of individuals to genotype by pooling DNA samples from a reference population.

Pooling of DNA samples and sequencing have provided a cost-effective alternative for a wide range of genomic applications, such as population genetics (Gautier et al., 2013), genome-wide association studies (Sham et al., 2002), and estimation of SNP effects for quantitative traits (Henshall et al., 2012; Bell et al., 2017). In a theoretical way, estimation of marker effects from a pooled DNA differs from standard individual genotypes in some aspects. First, in pooled DNA samples, only marker allele frequencies can be estimated, whereas in standard genotyping, individual marker genotypes are obtained. Second, in pooled DNA samples, marker allele

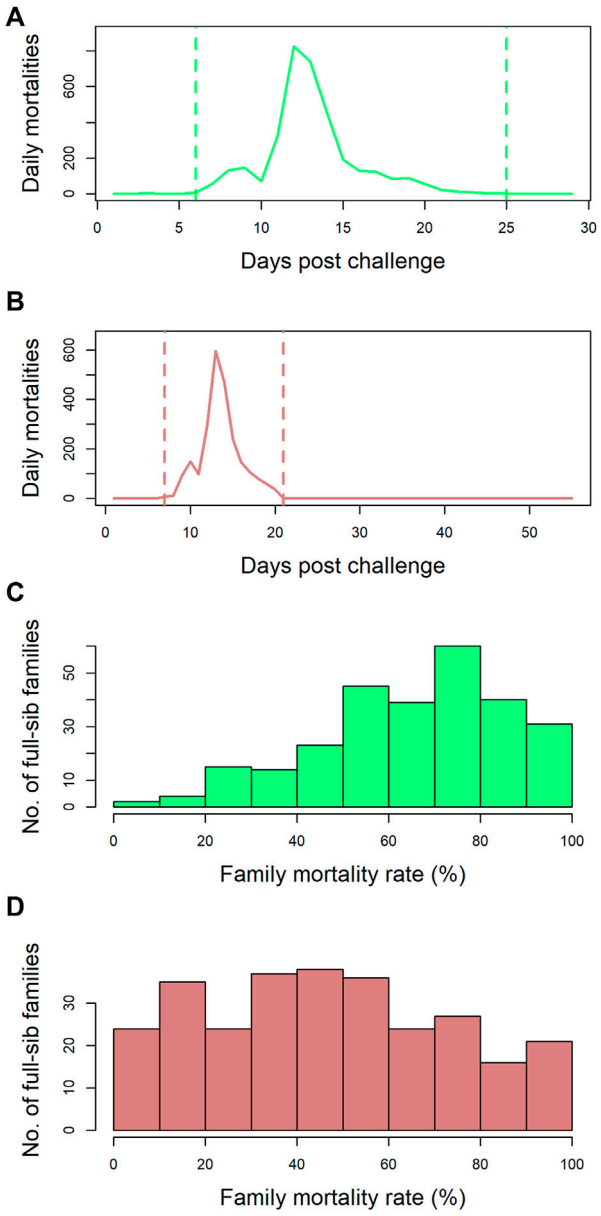

FIGURE 1
Mortality rate profiles of the datasets: **(A)** the number of mortalities observed per day over the course of the challenge trial (29 days) for dataset-1. Mortalities started 5 days post challenge and ended 25 days post challenge with the peak mortality observed at 12 days post challenge. **(B)** the number of mortalities observed per day over the course of the challenge trial (56 days) for dataset-2. Mortalities started 7 days post challenge and ended 21 days post challenge with the peak mortality observed at 13 days post challenge. **(C)** dataset-1—the number of full-sib families across the 273 full-sib families with a given percentage mortality. The mortality rate ranged from 0% to 100%, with an average mortality rate of 67%. **(D)** dataset-2—the number of full-sib families across the 282 full-sib families with a given percentage mortality. The mortality rate ranged from 0% to 100%, with an average mortality rate of 48%.

frequencies will normally be estimated with some degree of technical error, unequal contribution of sequenced reads derived among the individuals in a specific pool. Third, for DNA pools, the quantitative trait value of each individual cannot be assigned to a particular marker genotype, since information is not available on individual genotypes. However, by using the allelic frequencies at each tail to estimate the respective genotype frequencies and by assigning the sample average at each tail to every individual at that tail, the problems raised by the above differences can be overcome (Henshall et al., 2012; Gautier et al., 2013).

Sonesson et al. (2010) studied the use of DNA pooling of test individuals in combination with communal rearing of families as a means of reducing genotyping costs in aquaculture GS schemes using simulation. The study reported up to 0.88 accuracy of selection depending on the number of test individuals and the number of markers. However, to date, the potential of DNA pooling of a reference population for genomic prediction using data from practical breeding work is lacking. Therefore, in this study, we demonstrate the potential of DNA pooling for GS using both pooled DNA samples and *in-silico* DNA pooling. The effect of the number of pools and sequencing depth on the selection accuracy is studied. The study uses datasets generated for a breeding work to improve resistance against pancreas disease (PD).

2 Materials and methods

2.1 Datasets

Two datasets from the SalmoBreed breeding population of two year-classes (YC) were used for this study. Dataset-1 was from YC 2013, and dataset-2 was from YC 2015. The trait of interest was resistance to PD. PD is currently among the most economically important diseases in Norwegian Atlantic salmon production. It is caused by a salmonid alphavirus (SAV) of which SAV2 and SAV3 variants are found in Norway. These datasets were generated as part of the SalmoBreed's annual challenge test for the practical breeding work to improve host resistance against PD. The first dataset was used for real DNA sample pooling and the other was used for *in-silico* pooling.

2.1.1 Dataset-1: DNA pooling of samples

Data and tissue samples on 5,223 postsmolts belonging to 273 full-sib families from the SalmoBreed elite population challenged with SAV3 were available. The mortality profiles for this dataset are presented in Figures 1A and C. Mortalities started 5 days post challenge and ended 25 days post challenge with the peak mortality observed at 12 days post challenge (Figure 1A). A histogram of the number of full-sibling families across the 273 full-sib families with a given percentage of mortality is plotted in Figure 1C. The full-sibling mortality rate ranged from 0 to 100%, with an average mortality rate of 67% (Figure 1C). The individuals could be selected and grouped using survival information (early dead and/or late survivors) from the challenge test. Hence, 855 individuals were selected to develop four pools (i.e., M1, M2, S1, and S2 with the initial "M" representing the pool of mortalities/dead and "S" denoting the pool of surviving individuals).

DNA was extracted from each selected individual, quantified using the Quant-iT PicoGreen dsDNA assay kit, and normalized to a standard concentration for every individual; subsequently, equal quantities/volumes of DNA from each individual were pooled to make a specific pool group (M1, M2, S1, and S2). The M1 and M2 pools contained DNA representing 173 and 204 mortalities, respectively, while pools S1 and S2 incorporated DNA from 205 and 273 surviving individuals, respectively. The dead individuals of pools M1 and M2 represented 28 families, while pools S1 and S2 were represented by 35 and 34 families, respectively.

Libraries were prepared for sequencing using the Illumina PCR-free genomic DNA sample prep kits which were sequenced to approximately 40× depth. The sequencing was performed with an Illumina NextGen 500 instrument to obtain paired-end sequence reads of 150 bp. Trimmomatic software was used to perform adaptor and quality trimming of the generated sequence reads, and subsequently, high-quality sequence data were aligned to the Atlantic salmon genome reference sequence (assembly ICSASG_v2) using BWA-MEM version: 0.7.13-r1126 (Li, 2013). SNP detection, genotype calling, and allele frequencies on each locus were obtained using SAMtools version: 1.2. Software (Li et al., 2009). All the individuals pooled into four pools (M1, M2, S1, and S2) were also individually genotyped using ~57 K axiom Affymetrix SNP Genotyping Array (NOFSAL2). The overlapping SNPs across genotyping methods (sequencing vs. axiom array) were identified which yielded 45812 SNPs in common that were used for further genomic analyses.

2.1.2 Dataset-2: *In-silico* DNA pooling

Data from 4,115 postsmolts belonging to 282 full-sib families from the SalmoBreed elite population challenged with SAV3 were available. The mortality profiles for this dataset are presented in Figures 1B and D. Mortalities started 7 days post challenge and ended 21 days post challenge with peak mortality observed at 13 days post challenge (Figure 1B). A histogram of the number of full-sibling families across the 282 full-sib families with a given percentage of mortality is plotted in Figure 1D. The full-sibling mortality rate ranged from 0 to 100%, with an average mortality rate of 48% (Figure 1D). From the dataset, 914 individuals (435 survivors and 479 dead), belonging to 65 full-sib families, were selected based on family-wide mortality rates. The dead individuals represented 58 families, while the survived individuals represented 60 families. The data were split into a reference set (589 samples: 308 survivors and 281 dead) and a validation set

TABLE 1 Summary of the number of pools and the average number of individuals and families per pool per phenotype group.

# Pools	Dead pool		Alive pool	
	# Fish/pool	# Family/pool	# Fish/pool	# Family/pool
1	281	58	308	60
2	140.5	49.5	154	56
4	70.25	37.75	77	42.5
10	28.1	22.2	30.8	24.6
20	14.05	12.35	15.4	13.7
40	8.3	7.7	9.22	7.65
100	2.81	2.76	3.08	3.03
150	1.87	1.85	2.05	2.05
200	1.41	1.4	1.54	1.52

(325 samples: 127 survivors and 198 dead). Splitting of individuals into reference and validation sets was performed randomly within a full-sib family; hence, each family was represented in both datasets.

Genotypes for the customized NOFSAL2 SNP array with ~57 K SNPs were available for each individual. Pooled genotypes were generated in-silico as the frequency of alleles of the individuals in the pools and no random error was added to the pool genotypes.

2.2 Calculation of SNP allele frequencies

2.2.1 Allele frequency in DNA pools (dataset-1)

The pools were sequenced on an average sequencing depth of ~40× and the allele frequencies from each pool (S1, S2, M1, and M2) were obtained using a customized script implemented with *bcftools*, version: 1.10.2 (Danecek et al., 2021). The observed total allelic depth/frequency for each discovered variant in each pool was obtained using "*INFO/AD*" and the expression "*FORMAT/AD*" was used to obtain the observed frequency of the alternative allele. Moreover, the sequencing depth of each pool was reduced to 20× by random sampling of sequence reads followed by alignments, variants calling, and estimates of frequencies for both reference and nonreference alleles. The objective to reduce the sequencing depth per pool was to test the effect on the accuracy of prediction when the sequencing depth is reduced.

2.2.2 Allele frequency in *in-silico* DNA pooling (dataset-2)

The reference set was used for the calculation of allele frequencies by in-silico pooling of the genotypes of individuals in this set. They were pooled into different numbers of pools (i.e., 1, 2, 4, 10, 20, 40, 100, 150, and 200) per phenotype group (i.e., dead and survivors). Summaries of the average number of individuals and families per pool and per phenotype group are presented in Table 1. Allele frequencies from each pool were calculated by sampling with replacement given the individual genotypes. The number of times the sampling with replacement was performed is related to the average sequence depth, and it was done 20×, 40×, and 100×. This process was repeated independently 60 times.

2.3 Estimation of marker effects

The main difference between individual genotypes and pooled DNA regarding estimation of marker effects is that for the latter, a trait value of individuals cannot be assigned to a particular marker genotype exclusively. Therefore, marker effects are to be estimated from marker allele frequencies calculated from pooled DNA.

For the individual genotypes, marker effects were estimated by fitting the marker-based genomic model (SNP-BLUP):

$$y_{ind} = 1\mu + Xb_{ind} + e$$

where y_{ind} is the vector of phenotypes for the trait, μ is the overall mean, 1 is the vector of 1s, X is the matrix of genotypes dosage for all SNP coded as 0,1, and 2 and for all animals, b_{ind} is the vector of marker effects and it is assumed that $b_{ind} \sim N(0, I\sigma_m^2)$ and e is the vector of random residual and assumed $e \sim N(0, I\sigma_e^2)$ where σ_m^2 is the variance of marker effects, I is the identity matrix, and σ_e^2 is the residual error variance.

For the pools, the marker effects are estimated by fitting a slightly modified marker-based genomic model:

$$y_{pool} = 1\mu + Zb_{pool} + e_{pool}$$

where y_{pool} is the vector of phenotypes of the pools for the trait, μ is the overall mean, 1 is the vector of 1s, Z is the matrix of average allele frequencies for all SNPs and for all pools, b_{pool} is the vector of marker effects estimated from allele frequencies of pools, and e_{pool} is the vector of random residual. It is assumed that $b_{pool} \sim N(0, I\sigma^2_{m-pool})$ and $e_{pool} \sim N(0, I\sigma^2_{e-pool})$, where σ^2_{m-pool} is the variance of marker effects estimated from the marker frequencies of the DNA pools and σ^2_{e-pool} is the residual error variance. The analyses for both models were done using singular value decomposition based SNP-BLUP (Ødegård et al., 2018), which is suitable for large-scale genomic predictions.

2.4 Estimation of genomic breeding values

Estimation of genomic breeding values (GEBVs) for the selection candidates was performed by summing the effects of the markers multiplied by the standardized genotypes. Two GEBV values per individual were predicted using SNP effects estimated using pool data and the individual genotypes.

$$GEBV_j = \sum_i^m X_j b_i$$

where $GEBV_j$ is the vector of predicted GEBVs for individual j, X_i is the standardized genotypes for individual j, and b_i is the calculated marker effects either from pools of samples or individual genotypes.

2.5 Model evaluation

For dataset-1 DNA pooling of samples, the effect of sequencing depth on the accuracy of selection was studied by varying the sampling from originally available 40× to 20×. SNP effects from the pool data were calculated based on allele frequencies estimated from the pools and used to calculate GEBVs as described in the estimation of GEBV section.

For dataset-2 in-silico DNA pooling, the effect of sequence coverage on the accuracy of selection was studied by varying the sampling times to 40× and 100×. SNP effects from the pool data were calculated based on allele frequencies estimated from the pools and compared with SNP effects estimated from individual genotypes.

The accuracy of selection was calculated as the correlation between predicted GEBVs and phenotypes and weighted by the square root of the heritability (h^2 = 0.3).

$$Accuracy\ (r_{corr}) = \frac{\rho(GEBV, y)}{\sqrt{h^2}}$$

Where ρ is the Pearson moment correlation coefficient, $GEBV$ is the estimated GEBVs, y is the adjusted phenotype, and h^2 is the heritability of the trait.

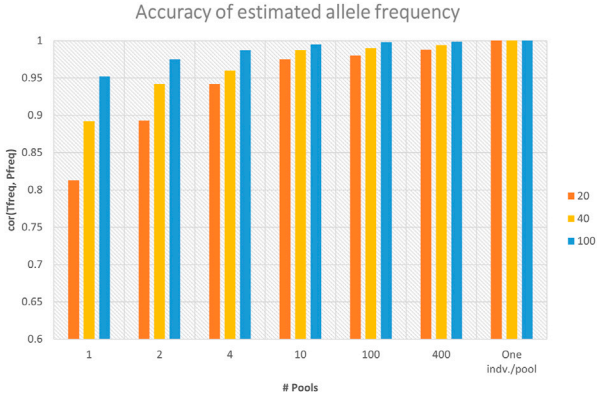

FIGURE 2
Accuracy of marker allele frequency calculated from different numbers of in-silico DNA pools and sequencing depths. Accuracy was calculated as the Pearson correlation coefficient between the true allele frequencies ("Tfreq") (i.e., calculated from the individual genotypes) and frequencies calculated using in-silico DNA pools ("Pfreq") (n = 1, 2, 4, 10, 100, and 914). There were 914 individuals, and the maximum number of pools (n = 914) represents one individual per pool.

3 Results

3.1 Accuracy of allele frequency estimation

Sequencing of DNA pools from individuals gives estimates of allele frequencies at SNPs. For dataset-1, the observed number of alleles at each locus for the sequenced pools provided the estimates of allele frequencies. The accuracies of marker allele frequency calculated from the in-silico DNA pools using three sequencing depths (i.e., 20×, 40×, and 100×) are presented in Figure 2. The accuracies were calculated as the Pearson correlation coefficients between the true allele frequencies (i.e., calculated from the individual genotypes) and frequencies calculated using in-silico pools. The figure shows that the accuracy of allele frequency estimation is affected by the number of pools and the average sequence depth coverage. As the number of pools increased from one pool, where all individuals represent one pool, to the maximum number of pools where each individual denotes a pool, the accuracy of allele frequency calculation has improved significantly, especially with low average sequence coverage (Figure 2). For example, for the sequencing coverage 40×, the Pearson correlation coefficients between DNA pool- and individual-based allele frequency estimation increased from 0.892 when only one DNA pool is considered to 0.99 when 10 DNA pools are used (Figure 2). Similar trends were also observed for the different sequencing coverages (Figure 2). As can be also seen from the figure, increasing sequence coverage

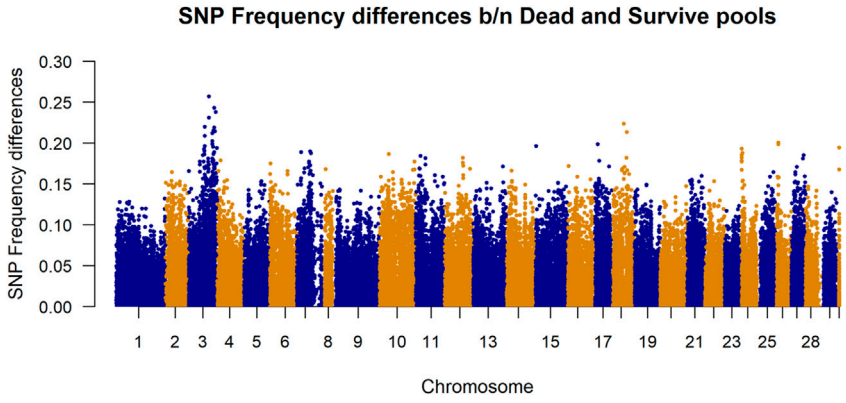

FIGURE 3
Manhattan plot of allele frequency differences between alive and dead pool for each SNP. The plotted allele frequency differences were calculated from 100 pools per phenotype group. Chromosome 30 represents markers belonging to unknown chromosome(s).

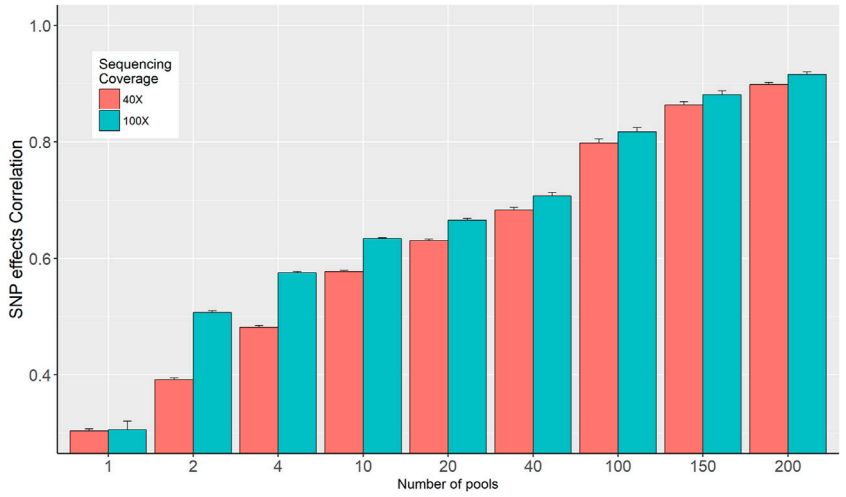

FIGURE 4
Correlation between SNP effects estimated from individual genotypes and in-silico DNA pool genotypes.

depth also improved the accuracy of allele frequency estimation, particularly when the number of pools is small (Figure 2). When a single pool per individual was used (i.e., no. of pools = 914), the correlation between allele frequencies calculated using individual genotypes and the pool is equal to 1, regardless of the sequencing coverage.

Differences in SNP frequencies between survivor and dead in-silico DNA pools are presented in Figure 3. The plotted allele frequency differences in the figure are absolute values of the frequency differences calculated from 100 pools per phenotype group. It shows that there are larger differences in frequencies between dead and survivor pools for the SNPs located at chromosome 3.

3.2 Correlation between SNP effects

The correlations between SNP effects estimated from individual genotypes and based on allele frequencies calculated from the in-silico DNA pools are presented in Figure 4. The correlation increased from 0.3 to 0.898 and from 0.311 to 0.912 for the 40× and 100× sequence coverage respectively, when the number of pools increased from 1 to 200 per phenotype group (Figure 4). Overall, it was observed that the impact of sequencing coverage on the correlation between SNP effects is limited; however, its importance increases when the number of pools is decreasing. Exception from the general trend was observed when 1 pool per phenotype group was used,

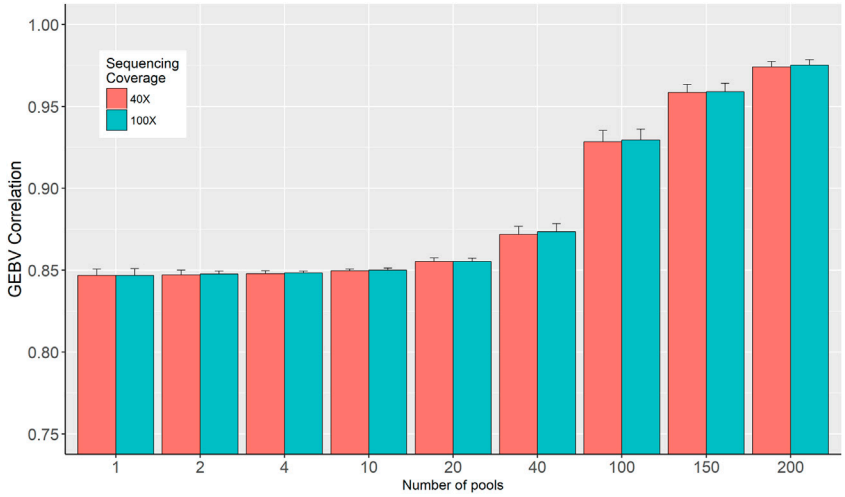

FIGURE 5
Correlation between genomic breeding values (GEBV) estimated from individual genotypes and in-silico DNA pool genotypes.

TABLE 2 Accuracy of selection of dataset-2–in-silico DNA pooling.

No. of pools	Accuracy of prediction			% Decreased	
	Individual	40×	100×	40×	100×
1	0.712 ± 0.005	0.574 ± 0.006	0.575 ± 0.006	19.38	19.24
2		0.574 ± 0.004	0.575 ± 0.002	19.382	19.24
4		0.575 ± 0.003	0.575 ± 0.002	19.242	19.24
10		0.576 ± 0.003	0.576 ± 0.001	19.101	19.10
20		0.581 ± 0.005	0.581 ± 0.002	18.54	18.40
40		0.594 ± 0.008	0.596 ± 0.006	16.57	16.29
100		0.642 ± 0.009	0.641 ± 0.011	9.83	9.97
150		0.667 ± 0.012	0.671 ± 0.011	6.32	5.76
200		0.684 ± 0.012	0.687 ± 0.012	3.93	3.51

Accuracy of prediction using *in-silico* DNA pools of the reference population for different numbers of pools and sequencing coverage. The % decreased is the decrease in accuracy of prediction in % for the 40× and 100× compared to the individual genotype. The presented accuracies are the mean of 60 replicates and the standard errors are the standard deviation of 60 replicates.

where no difference in SNP effects correlation was observed for 40× and 100× (Figure 4).

3.3 Correlation between genomic breeding values

Two sets of GEBVs for the validation dataset were calculated based on SNP effects estimated from individual genotypes and in-silico pooled DNA. Pearson correlation coefficients between these two sets of GEBVs are presented in Figure 5. The figure shows that the number of pools is the determining factor as the correlation increased from 0.84, when only a single pool is used per phenotype, to 0.976 when the number of pools increased to 200 for 40× sequencing coverage (Figure 5). These correlations are barely affected by the sequence coverage.

3.4 Genomic prediction accuracy

For the validation dataset in dataset-2–in-silico DNA pooling, genomic prediction accuracies were calculated as the Pearson correlation coefficients between true phenotypes and predicted GEBVs and weighted by the inverse of the square root

TABLE 3 Accuracy of prediction for dataset-1–DNA pooling of samples.

Dataset	No. of SNPs	Accuracy of prediction
SNP chip	51646	0.737 ± 0.006
Sequence 40×	44538	0.716 ± 0.004
Sequence 20x	45812	0.700 ± 0.003

of heritability of PD (Table 2). The result showed that the prediction accuracy for the individual genotypes was 0.712 and for the in-silico DNA pools, it ranged from 0.574 to 0.687 when the number of pools increased from 1 to 100. Table 2 also presents % decreased, the decline in accuracy for the DNA pools compared to the individual genotypes. Regardless of the sequencing coverage, approximately a 20% decline in accuracy was observed for in-silico DNA pools when less than 10 pools per group were used. However, the loss in accuracy was reduced to less than 10% when 100 pools per phenotype group were used. Furthermore, less than a 4% loss in accuracy was observed when up to 200 pools were used. The difference between 40× and 100× sequencing coverage with respect to prediction accuracy was limited (Table 2).

The accuracy of prediction values for dataset-1–DNA pooling of samples and individual SNP chips are shown in Table 3. The accuracy of prediction was 0.737 for the individual SNP chip data, and 0.716 and 0.700 for pooled data when sequencing coverage was 40× and 20×, respectively. This is up to ~5% higher accuracy for the individual SNP chip data.

4 Discussion

Implementation of a conventional GS in aquaculture species is very expensive because of the very large number of selection candidates and test-sibs to be genotyped. In recent years, cost-efficient GS design approaches either to minimize the number of individuals or the number of markers to genotype, without significantly reducing the accuracy of selection, have been given emphasis (Lillehammer et al., 2013; Dagnachew and Meuwissen, 2019; Tsairidou et al., 2020). This study investigates the potential of DNA pooling for creating a reference population for genomic prediction of PD resistance (binary observation) in salmon. It demonstrated that SNP effects in a reference population can be estimated from SNP allele frequencies that are calculated from DNA pools and then the GEBVs for selection candidates can be economically computed with acceptable accuracies.

The datasets used in this study were generated as part of the SalmoBreed's practical breeding work to improve host resistance against PD. PD is currently one of the most economically important diseases in the Norwegian production of Atlantic salmon (Jansen et al., 2010). The disease is caused by SAV, of which at least three distinct genotypes have been identified (Hodneland et al., 2005; Fringuelli et al., 2008). Mortality rates vary widely from PD outbreaks; survivors may eventually die because of secondary infections and increased parasitism and suffer reduced growth and degraded product quality (Lerfall et al., 2012). Moderate heritabilities have been reported from field outbreak data (Norris et al., 2008) and through controlled challenge testing using different challenge models (Gonen et al., 2015). Considerable efforts have been made in mapping genes for PD resistance for use in marker-assisted selection and a QTL for PD resistance was mapped to Atlantic salmon chromosome 3, using both fry and smolt challenge test data from two populations (Gonen et al., 2015; Hillestad et al., 2020). The difference in SNP frequencies between survivors and dead in-silico DNA pools (Figure 3) shows that there are larger differences in frequencies between dead and survivor pools for the SNPs located at chromosome 3. This validates that SNPs in that region are associated with some biological mechanisms which are likely to be influencing resistance to PD.

A sequence of DNA pools from individuals gives estimates of allele frequencies at SNPs with small or no loss in accuracy for a considerably lower cost compared to individual genotyping. Estimation of marker effects relies heavily on the accuracy of allele frequencies calculated from DNA pools. The accuracy of allele frequency estimation from pooled DNA samples depends on some experimental design parameters (Gautier et al., 2013), such as the number of individuals merged in a pool, the sequencing coverage, and the possibility of unequal contribution of each individual genome to the final sequencing read. The effects of the number of individuals merged in a pool were studied by varying the number of individuals in the pools (Table 1). However, the effect of sequencing coverage was studied by changing the number of sampling times (i.e., average sequencing coverage) for the in-silico pools and varying the sequencing coverage for pooled DNA samples. However, the effect of unequal contribution of individuals is not assessed in this study. It is important to note that accurate equimolar pooling of each genomic DNA is important for equal distribution of reads (Konczal et al., 2014) and the number of pooled samples should be balanced for accurate allele frequency estimation (Gautier et al., 2013; Konczal et al., 2014) and consequently, the implication of unequal DNA contribution for genomic prediction accuracy should be investigated.

For allele frequency calculation and SNP effects estimation, the importance of sequencing coverage decreased as the number of DNA pools increased (Figures 2 and 4). Given a fixed number of samples, as the number of pools increased, the number of individuals per pool reduced (Table 1), and thus the effectiveness of high sequencing coverage has diminished. This observed pattern is in agreement with that of Rellstab et al. (Rellstab

et al., 2013), who reported that higher sequencing coverages (>50) have no significant effect on allele estimation accuracy and only very low coverages (below 20×) would substantially reduce the precision. In the current study, an exception from the general trend was when 1 pool per phenotype group was used, where the advantage of sequencing coverage was visible for allele frequency accuracy (Figure 2) but not for the correlation of SNP effects (Figure 4). Furthermore, the SNP effect correlations were poor for both 40× and 100× coverage.

The accuracy of GS is expected to increase as the number of genotyped and phenotyped animals in the reference population increases for any trait, in particular, for lowly heritable traits (Solberg et al., 2008). Our results showed that SNP effects correlation, GEBV correlations, and prediction accuracy increased as the number of pools increased from 1 to 200 per phenotype group (Figures 4 and 5; Table 2). Henshall et al. (012) showed that a large number of smaller pools would estimate allele frequency more accurately than small numbers of large pools. As the number of pools increased, the accuracy of allele frequency estimation increased (Figure 2) and the prediction accuracy also increased (Table 2). Moreover, for a larger number of pools, it is observed that there are very little or no differences in SNP frequency accuracies and SNP effects correlation among different sequencing coverages (Figures 2 and 4). Similar trends were reported (Henshall et al., 2012; Bell et al., 2017) when there are a large number of pools. Furthermore, the high correlations between GEBVs estimated from individual genotype and DNA pools, particularly for the large number of pools (Figure 5), evidenced that there is limited to no reranking of individuals.

The accuracy of breeding values using pedigree information for dataset-2 (only using phenotypes of 914 individuals and their pedigree) was 0.48 (the result is not presented). This accuracy was significantly improved by the use of genomic information from the pools, which is in agreement with other reports (Sonesson et al., 2010; Dagnachew and Meuwissen, 2019; Kriaridou et al., 2020; Tsairidou et al., 2020), especially for the large number of pools. The loss of accuracy for the use of DNA pools compared with the individual genotypes was minimal for dataset-1 (0.74 vs. 0.71, Table 3). However, the prediction accuracy difference was substantial for dataset-2 (0.71 vs. 0.57, Table 2) for the same number of pools. One explanation is that for dataset-1, the prediction accuracies were obtained for the same individuals in the pools (i.e., reference and validation were the same individuals). On the other hand, for dataset-2, the validation individuals were different from the reference population. Furthermore, computer simulation of DNA pooling provides an approximation and might fail to capture some parameters.

Results from using DNA pooling for genomic prediction are lacking. Our results show a trade-off between the number of DNA pools and the loss of prediction accuracy. A reduction in prediction accuracy means a reduction in genetic gain. This has a cost implication that is complex to quantify as it is determined by the trait, the breeding goal, and other specifics of the breeding industry. Assuming $20 genotyping cost per individual and $300 cost of sequencing per sample for a 40× sequencing depth, the annual genotyping cost for 5,000 individuals (1,000 candidates and 4,000 informant sibs) is $100,000. However, for the DNA pool scenario, where only 1,000 candidates are genotyped and the reference siblings are pooled, the cost of genotyping varies from $20,600 to $140,000 for a single pool per phenotype to 200 pools per phenotype, respectively. Increasing the number of pools also increases the cost of sequencing and hence an appropriate pooling strategy should strike an optimal balance between cost-effectiveness and accuracy. Furthermore, in the present study, we have presented the prediction of allele frequencies from the pools of DNA using sequencing of the pools. However, calculation of allele frequencies from pooled DNA does not necessarily require sequencing of the pools. It has been reported that allele frequencies from DNA pools can also be calculated by SNP genotyping of the pools using light intensities (Reverter et al., 2014); hence, the cost associated with sequencing of the pools can be avoided.

As it is presented in the study, DNA pooling of a reference population can serve as a cost-effective GS approach, but with a potential limitation in that the identity of individuals would be lost and therefore individual characteristics and environmental factors could not be adjusted in genomic modeling, which may result in a loss in accuracy and a biased estimate of a genetic effect. In the current study, pooling within phenotype groups was done randomly; however, Henshall et al. (2012) suggested that pooling strategies within contemporary groups and fitting contemporary group in the model would eliminate some of these limitations. For example, in the studied datasets, pooling within sex and full-sib families would address these limitations.

5 Conclusion

DNA pooling of a reference population can serve as a cost-effective GS approach, but with some potential limitations. Results showed that a large number of pools are required to achieve as good genomic prediction accuracies as individual genotypes; however, alternative effective pooling strategies should be exploited to reduce the number of pools without reducing the prediction power. Nevertheless, it is demonstrated that pooling of a reference population can be used as a tool to optimize between cost and accuracy of selection.

Data availability statement

The datasets presented in this article are not readily available because it is co-owned by a third party. Request to access the datasets should be directed to anna.sonesson@nofima.no and hooman.moghadam@bmkgenetics.com.

Ethics statement

Ethical review and approval were not required for the animal study because the datasets used in this study were generated as part of Benchmark's routine breeding work, which is supervised and approved by the Norwegian Food Safety Authority (Mattilsynet).

Author contributions

BD: Conceptualization, investigation, methodology, writing—original draft, and writing—review and editing. MA: investigation, methodology, and writing—review and editing. BH: data curation and writing—review and editing. TM: methodology and writing—review and editing. AS: conceptualization, funding acquisition, investigation, methodology, and writing—review and editing.

Funding

This work received funding from the European Union's Seventh Framework Programme (KBBE.2013.1.2-10) under grant agreement no. 613611 (FISHBOOST). The funding body played no role in the design of the study, collection, analysis, and interpretation of data, and in writing the manuscript.

Conflict of interest

BH was employed by Benchmark Genetics.

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

References

Bell, A. M., Henshall, J. M., Porto-Neto, L. R., Dominik, S., McCulloch, R., Kijas, J., et al. (2017). Estimating the genetic merit of sires by using pooled DNA from progeny of undetermined pedigree. *Genet. Sel. Evol.* 49 (1), 28. doi:10.1186/s12711-017-0303-8

Dagnachew, B., and Meuwissen, T. (2019). Accuracy of within-family multi-trait genomic selection models in a sib-based aquaculture breeding scheme. *Aquaculture* 505, 27–33. doi:10.1016/j.aquaculture.2019.02.036

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 (2), giab008. doi:10.1093/gigascience/giab008

Fringuelli, E., Rowley, H. M., Wilson, J. C., Hunter, R., Rodger, H., Graham, D. A., et al. (2008). Phylogenetic analyses and molecular epidemiology of European salmonid alphaviruses (SAV) based on partial E2 and nsP3 gene nucleotide sequences. *J. Fish. Dis.* 31 (11), 811–823. doi:10.1111/j.1365-2761.2008.00944.x

Gautier, M., Foucaud, J., Gharbi, K., Cezard, T., Galan, M., Loiseau, A., et al. (2013). Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol. Ecol.* 22 (14), 3766–3779. doi:10.1111/mec.12360

Goddard, M. E., and Hayes, B. J. (2007). Genomic selection. *J. Anim. Breed. Genet.* 124 (6), 323–330. doi:10.1111/j.1439-0388.2007.00702.x

Gonen, S., Baranski, M., Thorland, I., Norris, A., Grove, H., Arnesen, P., et al. (2015). Mapping and validation of a major QTL affecting resistance to pancreas disease (Salmonid alphavirus) in Atlantic salmon (*Salmo salar*). *Heredity* 115 (5), 405–414. doi:10.1038/hdy.2015.37

Henshall, J. M., Hawken, R. J., Dominik, S., and Barendse, W. (2012). Estimating the effect of SNP genotype on quantitative traits from pooled DNA samples. *Genet. Sel. Evol.* 44, 12. doi:10.1186/1297-9686-44-12

Hillestad, B., Makvandi-Nejad, S., Krasnov, A., and Moghadam, H. K. (2020). Identification of genetic loci associated with higher resistance to pancreas disease (PD) in Atlantic salmon (*Salmo salar* L.). *BMC genomics* 21 (1), 388. doi:10.1186/s12864-020-06788-4

Hodneland, K., Bratland, A., Christie, K. E., Endresen, C., and Nylund, A. (2005). New subtype of salmonid alphavirus (SAV), Togaviridae, from Atlantic salmon *Salmo salar* and rainbow trout *Oncorhynchus mykiss* in Norway. *Dis. Aquat. Organ.* 66 (2), 113–120. doi:10.3354/dao066113

Jansen, M. D., Taksdal, T., Wasmuth, M. A., Gjerset, B., Brun, E., Olsen, A. B., et al. (2010). Salmonid alphavirus (SAV) and pancreas disease (PD) in Atlantic salmon, *Salmo salar* L., in freshwater and seawater sites in Norway from 2006 to 2008. *J. Fish. Dis.* 33 (5), 391–402. doi:10.1111/j.1365-2761.2009.01131.x

Konczal, M., Koteja, P., Stuglik, M. T., Radwan, J., and Babik, W. (2014). Accuracy of allele frequency estimation using pooled RNA-Seq. *Mol. Ecol. Resour.* 14 (2), 381–392. doi:10.1111/1755-0998.12186

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

Lerfall, J., Larsson, T., Birkeland, S., Taksdal, T., Dalgaard, P., Afanasyev, S., et al. (2012). Effect of pancreas disease (PD) on quality attributes of raw and smoked fillets of Atlantic salmon (*Salmo salar* L.). *Aquaculture* 324-325, 209–217. doi:10.1016/j.aquaculture.2011.11.003

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv arXiv:1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Lillehammer, M., Meuwissen, T. H. E., and Sonesson, A. K. (2013). A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. *Genet. Sel. Evol.* 45 (1), 39. doi:10.1186/1297-9686-45-39

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157 (4), 1819–1829. doi:10.1093/genetics/157.4.1819

Nielsen, H. M., Sonesson, A. K., Yazdi, H., and Meuwissen, T. H. E. (2009). Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquaculture* 289 (3–4), 259–264. doi:10.1016/j.aquaculture.2009.01.027

Norris, A., Foyle, L., and Ratcliff, J. (2008). Heritability of mortality in response to a natural pancreas disease (SPDV) challenge in Atlantic salmon, *Salmo salar* L., post-smolts on a West of Ireland sea site. *J. Fish. Dis.* 31 (12), 913–920. doi:10.1111/j.1365-2761.2008.00982.x

Ødegård, J., Indahl, U., Strandén, I., and Meuwissen, T. H. E. (2018). Large-scale genomic prediction using singular value decomposition of the genotype matrix. *Genet. Sel. Evol.* 50 (1), 6. doi:10.1186/s12711-018-0373-2

Ødegård, J., and Meuwissen, T. H. (2014). Identity-by-descent genomic selection using selective and sparse genotyping. *Genet. Sel. Evol.* 46 (1), 3. doi:10.1186/1297-9686-46-3

Rellstab, C., Zoller, S., Tedder, A., Gugerli, F., and Fischer, M. C. (2013). Validation of SNP allele frequencies determined by pooled next-generation sequencing in natural populations of a non-model plant species. *PLoS One* 8 (11), e80422. doi:10.1371/journal.pone.0080422

Reverter, A., Henshall, J. M., McCulloch, R., Sasazaki, S., Hawken, R., and Lehnert, S. A. (2014). Numerical analysis of intensity signals resulting from genotyping pooled DNA samples in beef cattle and broiler chicken. *J. Anim. Sci.* 92 (5), 1874–1885. doi:10.2527/jas.2013-7133

Sham, P., Bader, J. S., Craig, I., O'Donovan, M., and Owen, M. (2002). DNA pooling: a tool for large-scale association studies. *Nat. Rev. Genet.* 3, 862–871. doi:10.1038/nrg930

Solberg, T. R., Sonesson, A. K., Woolliams, J. A., and Meuwissen, T. H. E. (2008). Genomic selection using different marker types and densities. *J. Anim. Sci.* 86 (10), 2447–2454. doi:10.2527/jas.2007-0010

Sonesson, A. K., Meuwissen, T. H., and Goddard, M. E. (2010). The use of communal rearing of families and DNA pooling in aquaculture genomic selection schemes. *Genet. Sel. Evol.* 42, 41. doi:10.1186/1297-9686-42-41

Sonesson, A. K., and Meuwissen, T. H. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genet. Sel. Evol.* 41 (1), 37. doi:10.1186/1297-9686-41-37

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* 10 (2), 581–590. doi:10.1534/g3.119.400800

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership