scientific reports

Check for updates

OPEN

Transcriptome and resequencing reveal immune-related genes and molecular markers associated with *Aeromonas salmonicida* resistance in *Salmo trutta fario*

Jianshe Zhou[1], Shuaijie Sun[2], Wanliang Wang[1], Yaoqiong Zhang[1] & ZhuangZhuang Wang[1]✉

Understanding the molecular mechanisms underlying disease resistance in *Salmo trutta fario* is of fundamental importance for enhancing the health and sustainability of salmonid aquaculture. In this study, we performed an integrative multi-omics analysis combining transcriptome sequencing and whole-genome resequencing to systematically identify immune-related genes and single-nucleotide polymorphisms (SNPs) associated with resistance to *Aeromonas salmonicida* infection. Transcriptomic profiling identified multiple differentially expressed genes involved in immune and signaling pathways, including interferon-induced protein 44-like (IFI44L), PI3K family members (PIK3R2, PIK3R5), and NF-κB inhibitors (NFKBIA, NFKBIAA). These genes were significantly enriched in key immune regulatory pathways such as chemokine, PI3K-AKT, and NF-κB signaling, indicating their potential roles in modulating host defense responses. The integration of transcriptome and whole-genome resequencing data further revealed 104 SNPs distributed across immune-related genes that effectively differentiated resistant (R) and susceptible (S) populations. Among these, six candidate SNP loci were validated by PCR and demonstrated strong discriminatory power for resistant individuals, with multiplex PCR (MPCR) achieving an overall identification accuracy of 88.33%. The proportion of resistant genotypes within the population reached 14.82%, suggesting the progressive formation of a disease-resistant *S. trutta fario* strain. Overall, our findings provide novel and comprehensive insights into the genetic and transcriptional regulation of disease resistance in *S. trutta fario*. The identified immune-related genes and SNP markers represent valuable molecular resources for selective breeding and the development of health-oriented aquaculture strategies in salmonids.

Keywords *Salmo trutta fario*, *Aeromonas salmonicida*, Transcriptome, Genome resequencing, SNP, Disease resistance

The Qinghai–Xizang Plateau, known as the "roof of the world" and the "third pole of the Earth," plays a critical role in maintaining global ecological stability and biodiversity. Characterized by relatively limited anthropogenic disturbance, the region provides crucial water sources for major Asian rivers such as the Yangtze, Yarlung Zangbo, Lancang, and Nu. Its unique geographic and climatic features have contributed to the formation of a high degree of aquatic biodiversity[1]. However, the expansion of hydropower projects, particularly along the Yarlung Zangbo River, has increasingly intensified conflicts between resource development and aquatic ecosystem protection, making fish conservation an urgent ecological and socioeconomic concern[2]. *Salmo trutta fario*, belonging to the family Salmonidae, was introduced to the Yadong River in Shigatse, Xizang, during the mid-19th century[3,4]. It has since developed into a distinct aquaculture industry in the region. Among its major threats, *Aeromonas salmonicida*—the etiological agent of bacterial gill disease (BGD)—is particularly detrimental, leading to high mortality and substantial production losses[5,6]. Understanding the genetic mechanisms underlying resistance to *A. salmonicida* is therefore essential for the sustainable aquaculture development of *S. trutta fario.*

[1]Institute of Aquatic Sciences, Xizang Autonomous Region Academy of Agricultural and Animal Husbandry Sciences, Lasa 850032, China. [2]Henan Academy of Fishery Sciences, Zhengzhou 450012, China. ✉email: zwangzhuang@163.com

Transcriptome sequencing (RNA-seq) has become a cornerstone in functional genomics, enabling comprehensive and quantitative characterization of gene expression and uncovering molecular mechanisms underlying physiological responses and immune processes[7,8]. In parallel, whole-genome resequencing enables the detection of genome-wide variations—including single-nucleotide polymorphisms (SNPs), insertions/deletions (InDels), and structural variants—by comparison with a reference genome[9]. These technologies have been extensively applied to identifying genes linked to pathogen resistance and adaptive traits[10].

To elucidate the defense mechanisms of *S. trutta fario* against *A. salmonicida*, we first isolated and identified the causative agent of BGD through artificial infection and 16 S rRNA sequencing. Based on post-infection responses, fish were classified into resistant and susceptible groups. We then performed transcriptome sequencing on control (C), susceptible (S), and resistant (R) groups to identify differentially expressed genes (DEGs) and associated immune-related pathways. In parallel, whole-genome resequencing of liver tissue from S and R groups generated high-resolution data on genetic variation. Integrating transcriptomic and genomic analyses allowed us to pinpoint SNPs associated with resistance, followed by validation at the population level. This integrative approach provides mechanistic insights into host–pathogen interactions and offers practical molecular markers for disease-resistance breeding in *S. trutta fario*.

Materials and methods

Source and experimental design of fish

The *S. trutta fario* specimens utilized in this experiment were obtained from the Yarlung Zangbo River fishery resources breeding base in the Xizang Autonomous Region. A total of 360 healthy *S. trutta fario* were randomly selected for the experiment. At the beginning of the trial, the fish had an average body weight of 33.26 ± 8.11 g and an average total length of 14.81 ± 1.23 cm. Samples were classified into three experimental groups, each consisting of six *S. trutta fario* specimens. These groups were designated as follows: the control group (Group C, specimens labeled C1–C6), the susceptible group (Group S, specimens labeled S1–S6), and the resistant group (Group R, specimens labeled R1–R6). For whole-genome resequencing, samples were divided into two groups: the susceptible group (Group S) and the resistant group (Group R), each comprising six liver tissue samples from *S. trutta fario.*

Following experimental infection with *A. salmonicida*, fish were categorized into three groups based on survival status, clinical observations, and bacteriological examination. The susceptible group (S) comprised individuals who developed clear disease symptoms and/or died during the infection period, with *A. salmonicida* confirmed as the dominant bacterium isolated from internal organs. The resistant group (R) included individuals that survived the infection challenge without obvious clinical abnormalities, and in which no dominant proliferation of *A. salmonicida* was detected. The control group (C) consisted of fish injected with an equal volume of sterile physiological saline and exhibiting no abnormal clinical signs.

RNA and DNA extraction

A Total RNA was extracted from head-kidney tissue using a commercial RNA extraction kit (Tiangen Biotech, Beijing; protocol: https://www.tiangen.com/upload/file/20220826/20220826111922_59567.pdf). Genomic DNA was extracted following the manufacturer's instructions (https://www.tiangen.com/upload/file/20220826/20220826094350_61611.pdf). RNA and DNA integrity were verified using agarose gel electrophoresis and spectrophotometry. Only samples meeting quality control criteria were submitted to Shanghai Parsenor Biotechnology Co., Ltd. for RNA sequencing and whole-genome resequencing.

RNA-seq analysis and RT-qPCR validation

Sequencing was performed on the Illumina NovaSeq platform. Raw reads were processed using FastQC for quality assessment and Trimmomatic v0.39 for adapter trimming and removal of low-quality bases. Clean reads were aligned to the *S. trutta fario* reference genome (fSalTru1.1) using HISAT2 v2.2.1, and transcript assembly and expression quantification were performed with StringTie v2.1.5. Differential gene expression analysis was performed using DESeq2 v1.32.0. Genes with an absolute $\log_2$ fold change > 1 and an adjusted p-value (false discovery rate, FDR) < 0.05 were considered significantly differentially expressed. Before differential expression analysis, principal component analysis (PCA) and hierarchical clustering were conducted to evaluate sample relationships and detect potential batch effects. Gene Ontology (GO) and KEGG enrichment analyses were subsequently conducted to annotate gene functions and identify significantly enriched biological pathways. To validate the transcriptome sequencing results, RT-qPCR analysis was performed using six biological replicates ($n = 6$) for each experimental group. Total RNA was extracted from head kidney tissue using an RNA extraction kit (Tiangen Biotech, Beijing, China). RNA integrity was verified by 1.2% agarose gel electrophoresis, and RNA purity was assessed by measuring the OD260/280 ratio (2.01 ± 0.04). First-strand cDNA was synthesized using a reverse transcription kit (Baoyi Medical, Beijing, China). Gene-specific primers were designed based on differentially expressed gene sequences using NCBI Primer-BLAST, with amplicon sizes ranging from 73 to 163 bp. Primer specificity was confirmed by agarose gel electrophoresis and melting curve analysis. RT-qPCR reactions were performed in a 20 µL reaction mixture containing 10 µL SYBR qPCR Master Mix, 0.4 µL of each primer (10 µmol/L), 1 µL cDNA template, and nuclease-free water to volume. The amplification program consisted of an initial denaturation at 95 °C for 30 s, followed by 40 cycles of 95℃ or 10 s, 60 ℃ for 30 s, and 72 ℃ for 30 s. A melting curve analysis was performed at 95 ℃ for 10 s, 75 ℃ for 60 s, and 95 ℃ for 60 s to verify amplification specificity. The housekeeping genes hprt1 and rps29 were used as internal references. Relative gene expression levels were calculated using the 2^-ΔΔCt method. Each sample included three technical replicates, and results are presented as mean ± standard deviation.

Whole-genome resequencing and variant analysis

High-quality genomic DNA was used to construct sequencing libraries according to standard Illumina protocols. Libraries that passed quality control were sequenced on the BGI-T7 platform. Clean reads were aligned to the *S. trutta fario* reference genome using BWA-MEM v0.7.17, and PCR duplicates were removed with Picard v2.26.10. Variant calling was performed using GATK(v4.2.0.0) HaplotypeCaller[11]. SNPs and InDels were filtered based on read depth ≥ 10, mapping quality ≥ 30, and variant confidence (QUAL) ≥ 50, and were subsequently functionally annotated using SnpEff v5.0. GO enrichment analysis was performed using the topGO package based on the Gene Ontology (GO) database, with statistical significance evaluated using a hypergeometric test and enriched GO terms selected at $p < 0.05$, covering the three main GO categories: biological process (BP), cellular component (CC), and molecular function (MF). KEGG pathway enrichment analysis was conducted using the clusterProfiler package based on the KEGG database, with enrichment significance assessed using a hypergeometric test, and significantly enriched pathways were identified at $p < 0.05$, including both KO and pathway annotations.

Integration of transcriptome and resequencing data

SNP loci were compared between the S and R groups, with particular emphasis on SNPs located within differentially expressed genes (DEGs) identified by transcriptomic analysis with $p < 0.05$. Shared SNPs exhibiting distinct genotypic patterns between the two groups were selected as candidate molecular markers. SNP loci with clear genotype differentiation between the resistant and susceptible groups and located within differentially expressed genes were prioritized for downstream validation. Primer design for SNP validation was performed using the iGeneTech online primer design platform (https://m.primer.igenetech.com/) (Table 1).

Population-level validation of SNPs

A virulent *A. salmonicida* strain was cultured to a concentration of 1×10^8 CFU/mL. A total of 360 *S. trutta fario* were divided into three parallel groups (120 individuals per group) and injected intramuscularly with 0.1 mL of bacterial suspension. Based on post-infection survival outcomes and pathological observations, fish were classified into resistant (R) and susceptible (S) groups. Fin clips were preserved in ethanol for DNA extraction and MPCR-based validation of candidate SNPs.

Multiplex polymerase chain reaction (MPCR) was performed to validate selected SNP loci identified from transcriptome analysis. Genomic DNA was extracted from the fin tissue of *S. trutta fario* using a commercial genomic DNA extraction kit according to the manufacturer's instructions. DNA quality and concentration were assessed by agarose gel electrophoresis and spectrophotometric measurement before downstream analysis. Primers flanking the target SNP loci were designed using an online primer design platform based on the reference genomic sequences. Target regions were uploaded in BED format, and species-specific parameters were applied. Primer pairs with optimal ranking scores and minimal predicted secondary structures were selected for MPCR amplification. PCR products were subjected to Sanger sequencing (IGE Biotechnology Co., Ltd.). The obtained DNA sequences were aligned using SnapGene software, and SNP genotypes were determined by comparing sequencing chromatograms with reference sequences.

Gene name	Primer name	Primer sequence (5'-3')	Amplicon length	Gene ID
GAPDH	GAPDH-F	AGGCATCTCACAAACGAGGA	132	XM_029736584
	GAPDH-R	GGCAACAATCTCAACTCCCT		
RPL13a	RPL1I3a-F	CCTGTCTTTTACAGCGGGACA	142	XM_029730684
	RPL13a-R	AAGCAGAACTTGCTTCGCCA		
EF1a	EF1a-F	GCGCACAGTAACACCGAAAC	132	XM_029735647
	EF1a-R	GCCTOCGCACTTGTAGATCA		
ACTB	ACTB-F	GGTTCACTTTTGAGCCGTCAC	150	XM_029747716
	ACTB-R	CCGTTGTCAACAACCAGTGC		
HPRT1	HPRT1-F	TCCCAAAGCACTACGCTGAT	78	XM_029771658
	HPRT1-R	CTAGCCTCTCTGTCCTGTCAA		
β2M	B2M-F	GAGTACACCTGCAGAGTCCG	73	XM_029744303
	B2M-R	DAGATCCTTACATATTTGACTCCCA		
RPS29	RPS29-F	TCGGCTTTGTTAAGCTGGACT	86	XM_029733128
	RPS29-R	TTGCTGGATCATCCCGTTCC		
28 S rRNA	28 S rRNA-F	GTCCCTAACCACACAGTTTAACAT	141	XM_029746783
	28 S rRNA-R	CTGAGACATTTTATTGGCTCAAGGA		

Table 1. Information table of primer sequence of eight candidate reference genes. Eight commonly used housekeeping genes were initially selected as candidate reference genes. After a preliminary evaluation of expression stability across all samples, HPRT1 and RPS29 were identified as the most stable genes and were therefore used as reference genes for qRT-PCR normalization.

Animal welfare statement

The Ethical Committee approved all experimental procedures involving animals for Experimental Animals of the Xizang Autonomous Region Academy of Agricultural and Animal Sciences, Institute of Aquatic Sciences (Approval No. 2021005). All methods were carried out in accordance with the relevant guidelines and regulations. Before classifying fish as resistant or susceptible, pathological observations were conducted for a period of 14 days to monitor disease progression and ensure accurate group assignment. At the conclusion of the experiment, all fish were humanely euthanized in accordance with the AVMA Guidelines for the Euthanasia of Animals (2020) and the ARRIVE guidelines. Fish were first anesthetized in a buffered solution of tricaine methane sulfonate (MS-222; 150 mg/L) until loss of equilibrium and cessation of opercular movement were observed. After confirming deep anesthesia, euthanasia was completed using an overdose of MS-222 (300 mg/L) followed by pithing to ensure death. No fish regained consciousness during or after the procedure. All efforts were made to minimize animal suffering and stress throughout the study.

Statistical analysis

Statistical analyses were performed using R software (version 4.2.0). The association between SNP genotype patterns and resistance phenotype was evaluated using the chi-square test. When expected cell counts were less than five, Fisher's exact test was applied as an alternative. Classification accuracy of SNP-based prediction was calculated as the proportion of correctly classified individuals, and the 95% confidence interval (CI) was estimated using the binomial exact method. Cohen's kappa coefficient was calculated to assess the agreement between genotype-based classification and observed phenotypic resistance, and was interpreted according to standard guidelines (values of 0.61–0.80 indicating substantial agreement). A p-value < 0.05 was considered statistically significant.

Results

Statistical analysis of transcriptome sequencing data from the head kidney of *Salmo trutta fario*

Table 2 presents the descriptive statistics for transcriptome sequencing of 18 head kidney samples of *S. trutta fario*. Across these samples, a total of 875,102,466 raw reads were generated, yielding 800,275,148 clean reads after quality filtering. The average Q30 value for all samples was 93.15%, with each sample exceeding 90%, indicating high sequencing quality and reliability. The mean proportion of clean reads per sample was 91.46%, with all samples above 89%, reflecting effective data filtering and preprocessing. Regarding sequencing depth, the average number of clean reads was 42.04 million for Group C, 44.48 million for Group S, and 46.84 million for Group R.

Differential gene expression analysis

Transcriptome analysis identified 3,801 differentially expressed genes (DEGs) between Group C and Group S, including 983 upregulated and 2,818 downregulated genes. Comparison between Group S and Group R revealed

Sample	Reads number	Total number of bases	Q30 (bp)	N (%)	Q20 (%)	Q30 (%)	Clean reads number	Clean reads (%)
C1	50,589,240	7,638,975,240	7,175,071,576	0.000288	96.58	93.92	45,355,510	89.65
C2	44,021,214	6,647,203,314	6,214,245,718	0.000291	96.39	93.48	40,474,788	91.94
C3	46,824,910	7,070,561,410	6,581,919,180	0.000290	96.19	93.08	42,508,204	90.78
C4	46,480,944	7,018,622,544	6,544,743,872	0.000289	96.38	93.24	42,325,598	91.06
C5	44,655,036	6,742,910,436	6,280,244,113	0.000291	96.33	93.13	41,395,440	92.70
C6	43,815,280	6,616,107,280	6,153,687,293	0.000294	96.18	93.01	40,191,748	91.72
S1	48,220,626	7,281,314,526	6,812,147,736	0.000305	96.43	93.55	44,219,820	91.70
S2	45,255,190	6,833,533,690	6,380,978,679	0.000301	96.38	93.37	41,682,550	92.10
S3	44,766,634	6,759,761,734	6,295,928,393	0.000296	96.17	93.13	40,611,506	90.71
S4	55,232,394	8,284,859,100	7,654,600,365	0.003497	96.87	92.39	50,198,128	90.88
S5	45,863,212	6,925,345,012	6,481,793,341	0.000298	97.39	93.59	42,385,968	92.41
S6	52,383,798	7,909,953,498	7,389,308,120	0.000306	97.29	93.41	47,833,828	91.31
R1	55,564,802	8,390,285,102	7,875,038,716	0.000298	97.54	93.85	51,304,010	92.33
R2	50,913,370	7,687,918,870	6,979,417,208	0.000302	96.05	90.78	45,819,158	89.99
R3	49,346,952	7,451,389,752	6,934,495,955	0.000298	97.18	93.06	45,738,590	92.68
R4	50,793,008	7,669,744,208	7,139,763,014	0.000291	97.20	93.08	46,815,656	92.16
R5	54,250,408	8,191,811,608	7,643,429,639	0.000292	97.28	93.30	49,129,896	90.56
R6	46,125,448	6,964,942,648	6,504,846,369	0.000301	97.34	93.39	42,284,750	91.67
Total	875,102,466	132,085,239,972	123,041,659,287				800,275,148	
Average				0.000474	96.73	93.15		91.46

Table 2. Descriptive statistics of all 18 sample datasets for transcriptome sequencing in this study. Q30 (bp): The total number of bases with a base recognition accuracy of over 99.9%; N (%): The percentage of fuzzy bases; Q20 (%): The percentage of bases whose recognition accuracy is above 99%; Q30 (%): The percentage of bases with an accuracy of more than 99.9%.

2,528 DEGs, comprising 1,721 upregulated and 807 downregulated genes. Between Group C and Group R, 524 DEGs were detected, with 144 upregulated and 380 downregulated genes (Fig. 1A). The Venn diagram showed that 267 DEGs were shared between the Group C vs. R and Group C vs. S comparisons, while 99 DEGs overlapped between Group C vs. R and Group S vs. R. Notably, 1,607 DEGs were common to both Group C vs. S and Group S vs. R, and 30 DEGs were shared among all three comparisons (Fig. 1B). Volcano plot analysis indicated that although a greater number of genes were downregulated in Group C compared with Group S, the most statistically significant expression differences were observed among upregulated genes (Fig. 1C). Hierarchical clustering analysis revealed that samples from Group S and Group R clustered closely together, whereas Group C samples exhibited weaker clustering consistency (Fig. 1D).

Enrichment analysis of DEGs

GO enrichment analysis provided functional characterization of the identified DEGs (Fig. 2A). In the C vs. S comparison, DEGs were mainly enriched in cell membrane-related components (CC), cell surface receptor signaling and cell communication processes (BP), as well as iron ion binding and protein binding functions (MF). For the S vs. R comparison, DEGs were enriched in cell membrane components (CC), responses to system-level and immune-related processes (BP), and G protein-coupled chemokine receptor activity (MF). In the C vs. R comparison, DEGs were primarily enriched in the hemoglobin complex and cytoplasmic components (CC), gas and oxygen transport processes (BP), and oxygen and heme binding activities (MF). KEGG pathway enrichment analysis revealed that, in the C vs. S comparison, the most significantly enriched pathways included cytokine–cytokine receptor interaction, axon guidance, and transcriptional dysregulation in cancer. For the S vs. R comparison, cytokine–cytokine receptor interaction, cell adhesion molecules, and axon guidance were the dominant enriched pathways. In the C vs. R comparison, porphyrin and chlorophyll metabolism, IL-17 signaling, and other metabolic pathways were predominantly enriched (Fig. 2B). Across all comparisons, cancer-related pathways were consistently represented among the top 20 enriched pathways. Specifically, the C vs. S comparison involved 178 DEGs (48 upregulated and 130 downregulated), the S vs. R comparison involved 130 DEGs (80 upregulated and 50 downregulated), and the C vs. R comparison involved 34 DEGs.

Validation of differentially expressed genes by RT-qPCR

To validate the transcriptome sequencing results, three genes, representing an upregulated gene, a downregulated gene, and a non-differentially expressed gene, were randomly selected for expression analysis in head kidney tissue using RT-qPCR (Fig. 3). The expression patterns obtained by RT-qPCR were highly consistent with the transcriptome sequencing data, thereby confirming the accuracy and reliability of the RNA-seq results.

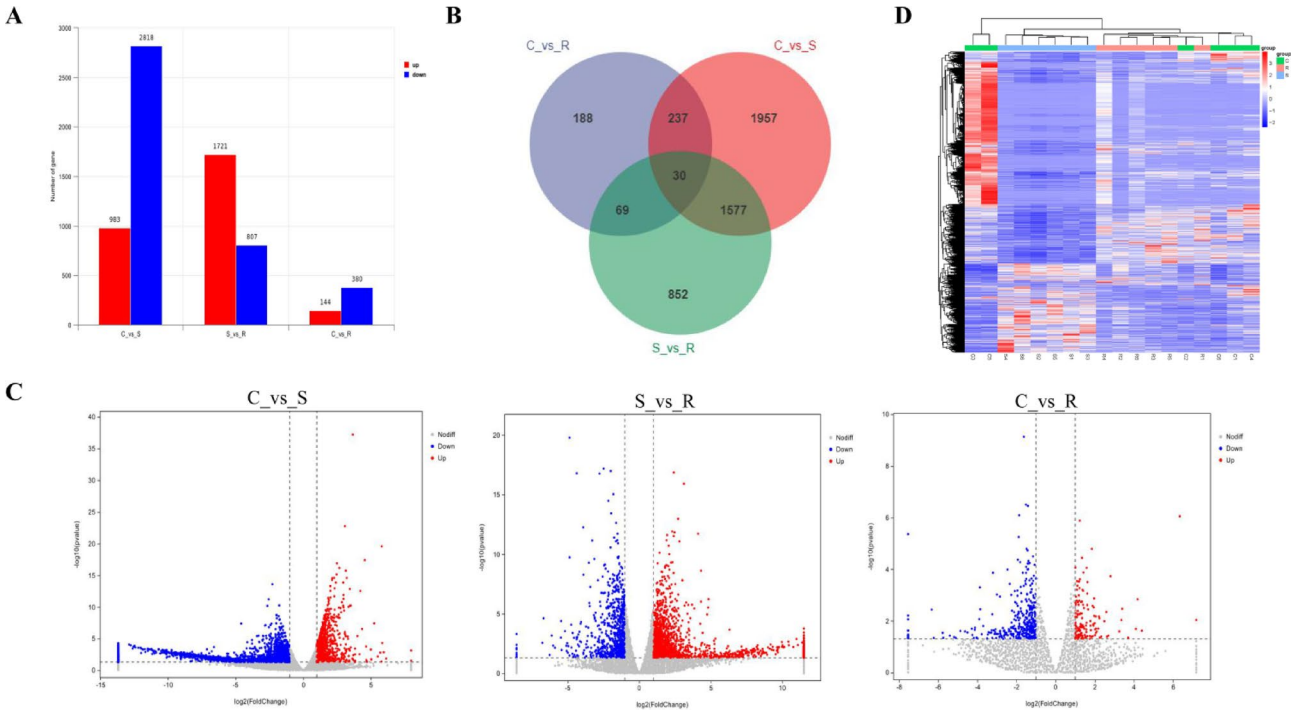

Fig. 1. Differentially expressed genes. (**A**) The number of differentially expressed genes. (**B**) Differential expression gene Venn map. (**C**) Heatmap of top 50 DEGs ($log_2FC > 1.5$, $p < 0.05$) between control ($n = 6$) and A. salmonicida-infected groups ($n = 6$). Color scale: blue = low expression ($log_2FPKM < 2$), red = high expression ($log_2FPKM > 4$). (**D**) Systematic clustering analysis of differentially expressed genes.

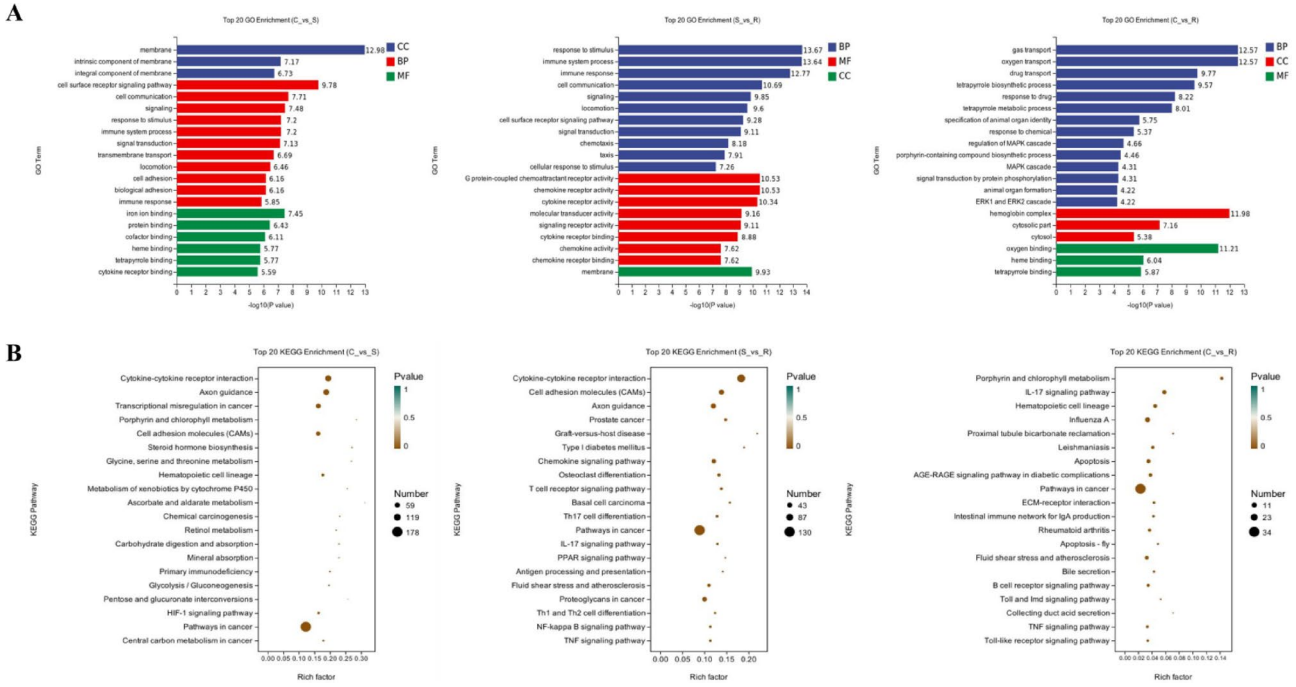

Fig. 2. Enrichment analysis of DEGs. (**A**) GO enrichment analysis histogram for pairwise comparison of three sets of samples. (**B**) The top 20 significant terms of KEGG enrichment.

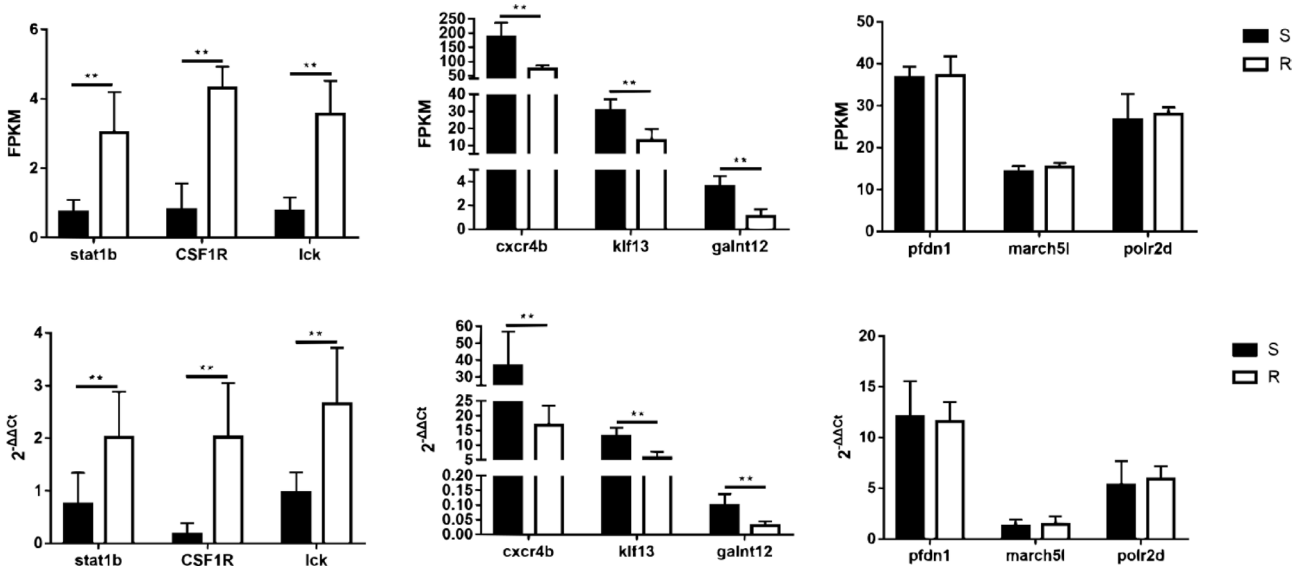

Fig. 3. Fluorescence quantitative PCR validation results.

Immune-related genes and pathway analysis

Based on transcriptome sequencing data (Table S1–S3), ten signaling pathways associated with organismal health, disease, and immune responses were identified. Among these pathways, the chemokine signaling pathway and the T cell receptor signaling pathway were particularly relevant to the immune responses of *S. trutta fario*. Within these two pathways, six differentially expressed genes (DEGs) were consistently detected: *pik3r2*, *pik3r5*, *nfkbiab*, *ENSSTUG0000007836*, *ENSSTUG00000003555*, and *nfkbiaa* (Table 3). These genes are involved in key immune-related functions, including T cell activation, cytokine-mediated signaling, and regulation of the NF-κB and PI3K signaling pathways. The identification of these DEGs highlights the molecular mechanisms underlying differential immune responses among the experimental groups and provides potential candidate targets for subsequent functional validation studies.

Rank	ID	FPKM			Foldchange (R/S)	*P*-value	*P*-adj	Regulation	Name	KEGG
		S	*R*	C						
1	ENSSTUG00000047153	51.53	108.05	112.13	2.10	0.000229	0.01271	Up	*pik3r2*	K02649
2	ENSSTUG00000046374	5219.92	1808.90	971.35	0.35	0.000410	0.01946	Down	*pik3r5*	K02649
3	ENSSTUG00000010218	1328.06	642.01	881.75	0.48	0.000482	0.02143	Down	*nfkbiab*	K04734
4	ENSSTUG00000007836	11.98	35.19	41.77	2.94	0.000572	0.02403	Up	-	K02649
5	ENSSTUG00000003555	2.14	13.28	9.86	6.20	0.001302	0.04232	Up	-	K07363
6	ENSSTUG00000040946	5962.07	2618.02	3340.92	0.44	0.001922	0.05552	Down	*nfkbiaa*	K04734

Table 3. List of genes shared in immune system-related pathways between groups S and R. FPKM values represent normalized expression levels used for visualization of gene expression across groups (S, R, and C), whereas differential expression statistics (Fold change, P-value, and adjusted P-value) were calculated using DESeq2 based on raw read counts.

Sample	Raw reads	Raw bases	Clean reads	Clean bases	Clean reads Q20 (%)	Clean reads Q30 (%)	Clean reads GC (%)
S1	168,400,252	25,260,037,800	168,104,814	24,964,358,406	94.5463	85.1868	43.5842
S2	169,762,716	25,464,407,400	169,473,354	25,160,950,094	94.6788	85.2571	43.4066
S3	172,842,952	25,926,442,800	172,508,308	25,613,556,838	94.5984	85.2140	43.5542
S4	170,081,068	25,512,160,200	169,767,426	25,216,349,880	94.5732	85.1485	43.4726
S5	167,268,848	25,090,327,200	166,915,318	24,793,968,670	94.5683	85.1464	43.4931
S6	192,185,584	28,827,837,600	191,826,140	28,501,101,652	94.7108	85.1902	43.1625
R1	185,830,770	27,874,615,500	185,558,426	27,557,681,350	94.6966	85.2626	43.3209
R2	165,659,920	24,848,988,000	165,323,654	24,564,546,172	94.5695	85.1586	43.3483
R3	163,877,556	24,581,633,400	163,715,762	24,336,220,370	95.3005	85.5450	43.1701
R4	162,323,222	24,348,483,300	162,129,566	24,114,648,758	95.4487	85.7697	43.0694
R5	183,388,130	27,508,219,500	183,062,840	27,198,206,844	94.6909	85.1782	43.3536
R6	176,709,630	26,506,444,500	176,413,382	26,203,189,168	94.7546	85.2310	43.1571
Total	2,078,330,648	311,749,597,200	2,074,798,990	308,224,778,202			
Average					94.7614	85.4407	43.3411

Table 4. Descriptive statistics of whole gene resequencing data for 12 samples.

Whole-genome resequencing and SNP screening

Whole-genome resequencing was performed on 12 liver tissue samples of *S. trutta fario*, generating 2,078,330,648 raw reads and 2,074,798,990 clean reads, with an average Q30 value of 85.44% (Table 4). After filtering, Group S exhibited 32,888 group-specific SNPs and 6,528 group-specific Indels, whereas Group R displayed 45,859 specific SNPs and 10,092 specific Indels. SNP distribution analysis revealed that variants were distributed across all 40 chromosomes (Fig. 4A), with the predominant mutation types being T: A > C: G and C: G > T: A transitions (Fig. 4B). Base preference analysis demonstrated a clear bias toward A and T bases at mutation sites (Fig. 4C).

By integrating transcriptome DEGs (2,528 genes identified between Groups S and R) with whole-genome SNP data, 104 SNP loci exhibiting distinct genotypic differences between the two groups were identified. These loci were distributed across 36 chromosomes, with no SNPs detected on chromosomes 15, 29, 34, and 36. Subsequent multiplex PCR validation using six randomly selected individuals per group confirmed six SNP loci that could consistently distinguish susceptible and resistant individuals (Table 5). This integrative strategy effectively combines transcriptomic and genomic variation data to identify candidate genetic markers associated with disease resistance.

Population validation of SNP markers

To further validate the identified SNP markers at the population level, 360 *S. trutta fario* individuals were artificially infected with *A. salmonicida*. Among them, 296 individuals (82.22%) developed severe clinical symptoms or died, and were classified as the susceptible group (S), whereas 64 individuals (17.78%) survived without obvious pathological symptoms and were classified as the resistant group (R). From each group, 30 individuals were randomly selected for multiplex PCR analysis (see Sect. 3.5). In the susceptible group, 26 individuals (85.67%) exhibited genotype combinations consistent with the susceptible profile across all loci, while in the resistant group, 27 individuals (90.00%) matched the resistant genotype pattern, resulting in an overall classification accuracy of 88.33% (95% CI: 80.22%-96.44%). The 53 individuals with genotype-phenotype concordance (26 + 27) represented 14.82% of the total population (53/360), reflecting the enrichment of disease-resistant genotypes. The association between SNP genotypes and resistance phenotype was statistically significant ($\chi^2 = 176.97$, $p < 0.001$), with a Cohen's kappa coefficient of 0.77, indicating substantial agreement between genotype and phenotype. These results demonstrate the robustness and reliability of the identified

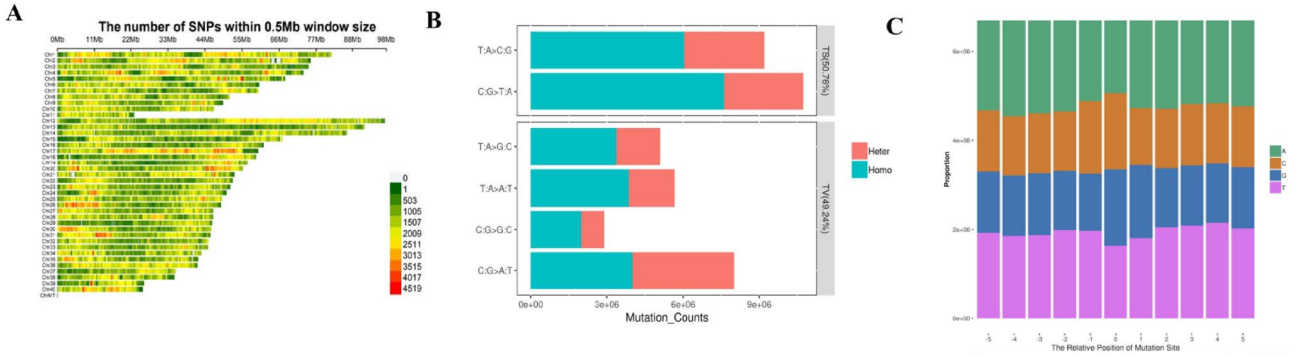

Fig. 4. Analysis of SNP Characteristics. (**A**) The distribution of SNPs on chromosomes. (**B**) SNP mutation map analysis. (**C**) Statistical distribution of base preference for single nucleotide polymorphism mutations. The horizontal axis represents the base position relative to the mutation site (0 indicates the mutation site; negative values represent bases upstream, and positive values represent bases downstream), while the vertical axis represents the proportion of different bases, reflecting base preference patterns around the mutation site.

Chromosome number	Locus	Ref	Alts	Susceptible genotype	Resistance genotype
27	11,029,129	T	G	G/G and T/T	T/G
4	16,842,638	G	A	A/A	G/A
12	77,517,128	C	T	C/C and T/T	C/T
13	87,581,629	A	G	A/G and G/G	A/A
18	18,173,793	T	C	T/T and T/C	C/C
40	7,014,825	C	T	T/T and C/C	C/T

Table 5. Six SNP loci information for whole genome resequencing combined with transcriptome screening.

SNP markers for distinguishing disease-resistant individuals and support their potential application in marker-assisted breeding programs for *S. trutta fario*.

Discussion

Transcriptome sequencing reveals differentially expressed genes and immune-related pathways

In this study, we investigated immune-related transcriptional responses in *S. trutta fario* following artificial infection with *A. salmonicida* and identified several key genes potentially involved in immune regulation and disease resistance. Notably, ENSSTUG00000036477, ENSSTUG00000018810, and ENSSTUG00000036537 were associated with interferon-induced protein 44-like (IFI44L), while aplnra was linked to the adipose inflammatory factor receptor A, and rrad to the GTP-binding protein Rad analog. IFI44 and IFI44L have been previously shown to play important roles in innate immunity and antiviral defense, with involvement in antigen presentation and NF-κB signaling pathways. Functional studies have demonstrated that IFI44L exhibits antiviral and antiproliferative activities and can serve as a diagnostic biomarker for viral infections[12,13]. The upregulated expression of IFI44L observed in our dataset, therefore suggests that this gene may contribute positively to resistance against *A. salmonicida* in *S. trutta fario*[14]. Moreover, the GTP-binding domain of IFI44 may be regulated by rrad, further indicating a potential cross-regulatory mechanism within interferon-mediated signaling pathways.

Transcriptome analysis also highlighted several immune-related signaling pathways, particularly the chemokine signaling pathway and the T cell receptor signaling cascade. Among the differentially expressed genes, *pik3r2* and *pik3r5* belong to the phosphoinositide 3-kinase (PI3K) family, which is known to modulate the PI3K/AKT signaling pathway. Expression of *pik3r2* has been associated with tumorigenesis and immune cell infiltration[15],, whereas inhibition of *pik3r5* has been reported to regulate AKT/mTOR signaling, epithelial–mesenchymal transition, and autophagy-related carcinogenesis[16]. The genes *nfkbiab* and *nfkbiaa*, members of the NF-κB inhibitor family, play critical roles in regulating NF-κB signaling, which functions as a central transcriptional regulator in both innate and adaptive immune responses[17]. Additionally, ENSSTUG0000007836 and ENSSTUG00000003555, both associated with PI3K-related signaling, may enhance T cell–mediated immune responses[18]. The PI3Kδ pathway is also known to regulate activation-induced cytidine deaminase (AID) expression in B cells, thereby influencing antibody diversification and adaptive immunity[19]. Collectively, these findings indicate that the immune response of *S. trutta fario* to *A. salmonicida* infection involves the coordinated regulation of multiple immune signaling pathways, providing important molecular insights into host defense mechanisms. Importantly, previous studies in salmonids have primarily reported activation of classical innate immune pathways, including interferon-mediated signaling, chemokine responses, and NF-κB

regulation, following *A. salmonicida* infection. Our results are consistent with these observations, confirming that interferon-related genes (e.g., IFI44L), PI3K signaling components (e.g., pik3r5), and NF-κB inhibitors (e.g., nfkbiaa) participate in antibacterial immune responses in salmonids. However, unlike earlier studies that mainly described transcriptional responses at the population level, the present study links these genes specifically to phenotypically defined resistant and susceptible individuals. This resistance-oriented analytical framework extends previous findings by proposing IFI44L-, PI3K-, and NF-κB–related genes as candidate markers associated with differential disease resistance in *S. trutta fario*. It is noteworthy that several cancer-related pathways were also significantly enriched in the KEGG analysis. These pathways share multiple core signaling components with immune and inflammatory processes, including PI3K–AKT, NF-κB, and MAPK signaling pathways. Therefore, their enrichment may reflect the activation of general cellular stress and immune regulatory mechanisms during bacterial infection rather than direct involvement in tumorigenesis. Consequently, candidate genes lacking clear functional annotation were interpreted cautiously in this study.

The relatively small sample size ($n = 6$ per group) represents a limitation of the present study and may reduce statistical power, potentially affecting the sensitivity of differential gene detection and the generalizability of transcriptomic inferences. Although larger cohorts (> 50 individuals per group) would enhance statistical robustness, similar biological replicate numbers (5–8 individuals per group) are commonly adopted in fish RNA-seq studies under controlled experimental designs. For instance, Gorgoglione et al. performed transcriptomic analyses in *S.trutta fario* using comparable sample sizes and successfully identified biologically meaningful immune-related gene expression patterns[20]. Nevertheless, future studies incorporating larger sample sizes and multi-population validation are warranted to further strengthen the robustness and applicability of these findings.

Whole-genome resequencing and transcriptome integration reveal candidate SNPs

To further elucidate the genetic basis of disease resistance, we integrated whole-genome resequencing and transcriptome data to identify single-nucleotide polymorphisms (SNPs) associated with resistance in *S. trutta fario*. Comparative analysis between resistant (R) and susceptible (S) groups identified 104 SNP loci distributed within differentially expressed genes. Previous studies have demonstrated the effectiveness of genotyping-by-sequencing (GBS) and reduced representation libraries (RRLs) for genome-wide SNP discovery in fish species, including the identification of 791 polymorphic SNPs in 190 blue catfish (*Ictalurus furcatus*) individuals[21,22]. Similarly, whole-genome resequencing of Han and Yanbian cattle populations has revealed genome-wide SNPs and selective signatures associated with important traits[23]. These studies highlight the importance of stringent filtering criteria and targeted screening strategies, including group-based selection and the integration of multi-omics datasets, to improve the accuracy and reliability of SNP identification. Integrating phenotypic information with genome-scale data has also proven valuable for identifying causal variants and disease-associated mechanisms. For example, combined whole-genome resequencing and transcriptome analyses have identified genes associated with oogenesis in autotetraploid crucian carp (*Carassius auratus*)[24]. In Atlantic salmon, genomic prediction models such as G-BLUP and Bayesian approaches have been successfully applied to enhance resistance to sea lice (*Caligus rogercresseyi*), achieving high prediction accuracy with approximately 10,000 SNP markers[25]. Comparable strategies in human genomics have linked expression quantitative trait loci (eQTLs) with GWAS data to identify causal variants influencing gene expression[26]. Building on these approaches, our study validated 104 candidate SNPs and further identified six loci capable of reliably distinguishing resistant and susceptible *S. trutta fario.* To our knowledge, SNP markers specifically associated with *A. salmonicida* resistance have not previously been reported in *S. trutta fario*. While genomic prediction models and SNP-based resistance studies have been successfully applied in other salmonids, our study provides the first integration of transcriptome-guided SNP discovery with phenotypic resistance classification in this subspecies. The six validated loci therefore represent newly proposed resistance-associated candidate markers, offering foundational resources for future marker-assisted selective breeding programs.

Multiplex PCR validation confirms the reliability of SNP markers

To verify the accuracy and practical applicability of the identified SNPs, we conducted population-level validation using multiplex PCR (MPCR). Since its initial development in 1988[27], MPCR has become a widely used and versatile technique for mutation and deletion[28,29], genotyping and quantitative analysis[30,31], and molecular diagnostics[32,33]. In the present study, six SNP-specific primer sets were designed and applied to juveniles from both the susceptible and resistant groups following *A. salmonicida* infection. Among the 30 susceptible individuals, 26 exhibited genotype combinations consistent with susceptibility across all six SNP loci, resulting in an accuracy of 85.67%. In the resistant group, 27 of 30 individuals displayed genotype profiles corresponding to resistance, achieving an accuracy of 90.00%. The overall prediction accuracy reached 88.33%. The proportion of resistant individuals in the total population was approximately 14.82%, indicating gradual enrichment of disease-resistant genotypes. These results demonstrate that MPCR-based SNP screening is an efficient and reliable method for identifying resistant individuals ($\chi^2 = 176.97$, $p < 0.001$; Cohen's kappa = 0.77), with a classification accuracy of 88.33% (95% CI: 80.22%-96.44%). Compared with previous salmonid studies that primarily relied on genome-wide association or genomic prediction models, our approach provides experimentally validated, locus-specific markers directly linked to resistance phenotype, thereby offering a practical and scalable tool for marker-assisted selection and health management in *S. trutta fario* aquaculture.

From a practical perspective, the SNP markers identified in this study provide promising tools for marker-assisted selection (MAS) in *S. trutta fario* breeding programs. By applying MPCR-based genotyping, large numbers of juveniles can be screened rapidly and non-invasively at early developmental stages without requiring pathogen challenge tests. This strategy enables efficient identification of individuals with potential resistance to *A. salmonicida*, facilitating the early establishment of disease-resistant populations in aquaculture systems.

Consequently, the application of these markers may help reduce disease outbreak risks, improve biosecurity in farming environments, and support the sustainable development of the *S. trutta fario* aquaculture industry.

Conclusion

Through the integrated analysis of transcriptome and whole-genome resequencing data, we identified multiple immune-related genes and key SNPs associated with resistance to *A. salmonicida* in *S. trutta fario*. The involvement of the chemokine signaling pathway and T cell receptor signaling pathway, together with critical genes such as *pik3r5*, *nfkbiab*, *nfkbiaa*, and *pik3r2*, highlights a conserved PI3K–NF-κB–mediated immune defense mechanism. A total of six SNP loci were confirmed as candidate molecular markers, with *pik3r5* and *nfkbiaa* showing the strongest associations with disease resistance. These findings advance our understanding of the genetic basis of disease resistance in cold-water salmonids and provide valuable molecular tools for marker-assisted selective breeding, supporting the sustainable development of *S. trutta fario* aquaculture on the Tibetan Plateau.

Data availability

The raw sequencing data generated in this study have been deposited in the Genome Sequence Archive (GSA) (Genomics, Proteomics & Bioinformatics, 2021) at the National Genomics Data Center (NGDC), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences. These datasets are publicly available under the accession numbers CRA036457 (https://download.cncb.ac.cn/gsa6/CRA036457) and CRA036448 (https://download.cncb.ac.cn/gsa6/CRA036448).

Received: 23 December 2025; Accepted: 16 March 2026
Published online: 25 March 2026

References

1. Zhang, C. et al. Status, problems and countermeasures of fishery resource conservation in Tibet. *Chin. J. Fisheries*. **27**, 68–72 (2014).
2. Zhou, J. S. et al. Research progress on fishery resources investigation in Tibet. *Chin. Agric. Sci. Bull*. **29**, 53–57 (2013).
3. Ferguson, A., Reed, T. E., Cross, T. F., McGinnity, P. & Prodöhl, P. A. Anadromy, potamodromy and residency in brown trout *salmo trutta*: The role of genes and the environment. *J. Fish Biol.* **95**, 692–718 (2019).
4. MacCrimmon, H. R., Marshall, T. L. & Gots, B. L. World distribution of brown trout, *salmo trutta*: Further observations. *J. Fish. Res. Bd Can.* **27**, 811–818 (1970).
5. Sun, S. et al. Selection of stable reference genes for RT–qPCR in salmo trutta. *Aquaculture Rep.* **26**, 101290 (2022).
6. Sun, S. J. et al. Isolation and identification of pathogenic bacteria causing bacterial gill disease in Salmo trutta. *Fish. Sci.* **42**, 288–295 (2023).
7. Kumar, G. & Kocour, M. Applications of next-generation sequencing in fisheries research: A review. *Fish. Res.* **186**, 11–22 (2017).
8. Wang, Z., Gerstein, M. & Snyder, M. RNA-seq: A revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
9. Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: Key considerations in genomic analyses. *Nat. Rev. Genet.* **15**, 121–132 (2014).
10. Yáñez, J. M. et al. High-throughput single nucleotide polymorphism (SNP) discovery and validation through whole-genome resequencing in nile tilapia (oreochromis niloticus). *Mar. Biotechnol.* **22**, 109–117 (2020).
11. McKenna, A. et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
12. Pan, H. et al. Interferon-induced protein 44 correlated with immune infiltration serves as a potential prognostic indicator in head and neck squamous cell carcinoma. *Front. Oncol.* **10**, 557157 (2020).
13. DeDiego, M. L., Nogales, A., Martinez-Sobrido, L. & Topham, D. J. Interferon-induced protein 44 interacts with cellular FK506-binding protein 5, negatively regulates host antiviral responses, and supports virus replication. *mBio* **10**, e01839–e01819 (2019).
14. Busse, D. C. et al. Interferon-induced protein 44 and interferon-induced protein 44-like restrict replication of respiratory syncytial virus. *J. Virol.* **94**, e00297–e00220 (2020).
15. Liu, Y. et al. Pan-cancer analysis on the role of PIK3R1 and PIK3R2 in human tumors. *Sci. Rep.* **12**, 5924 (2022).
16. Liu, W. et al. miR-210-5p promotes epithelial–mesenchymal transition by inhibiting PIK3R5 thereby activating oncogenic autophagy in osteosarcoma cells. *Cell. Death Dis.* **11**, 93 (2020).
17. Zhang, Q., Lenardo, M. J. & Baltimore, D. 30 years of NF-κB: A blossoming of relevance to human pathobiology. *Cell* **168**, 37–57 (2017).
18. Xu, K. et al. Glycolysis fuels phosphoinositide 3-kinase signaling to bolster T cell immunity.
19. Compagno, M. et al. Phosphatidylinositol 3-kinase δ blockade increases genomic instability in B cells. *Nature* **542**, 489–493 (2017).
20. Gorgoglione, B. et al. Comparative study of CXC chemokines modulation in brown trout salmo trutta following infection with a bacterial or viral pathogen. *Mol. Immunol.* **71**, 64–77 (2016).
21. Li, C. et al. SNP discovery in wild and domesticated populations of blue catfish, *I ctalurus furcatus*, using genotyping-by-sequencing and subsequent SNP validation. *Mol. Ecol. Resour.* **14**, 1261–1270 (2014).
22. Van Bers, N. E. M., Crooijmans, R. P. M. A., Groenen, M. A. M., Dibbits, B. W. & Komen, J. SNP marker detection and genotyping in tilapia. *Mol. Ecol. Resour.* **12**, 932–941 (2012).
23. Choi, J. W. et al. Whole-genome resequencing analysis of hanwoo and yanbian cattle to identify genome-wide SNPs and signatures of selection. *Mol. Cells*. **38**, 466–473 (2015).
24. Wang, C. et al. Whole-genome re-sequencing and transcriptome reveal oogenesis-related genes in autotetraploid carassius auratus. *Mar. Biotechnol.* **23**, 233–241 (2021).
25. Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P. & Yáñez, J. M. The use of genomic information increases the accuracy of breeding value predictions for sea louse (caligus rogercresseyi) resistance in atlantic salmon (salmo salar). *Genet. Sel. Evol.* **49**, 15 (2017).
26. The Geuvadis Consortium. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
27. Chamberlain, J. S., Gibbs, R. A., Rainer, J. E., Nguyen, P. N. & Thomas, C. Deletion screening of the duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucl. Acids Res.* **16**, 11141–11156 (1988).
28. Vossen, R. H. A. M., Aten, E., Roos, A. & Den Dunnen, J. T. High-resolution melting analysis (HRMA)-more than just sequence variant screening. *Hum. Mutat.* **30**, 860–866 (2009).

29. Zakrzewski, F. et al. Targeted capture-based NGS is superior to multiplex PCR-based NGS for hereditary BRCA1 and BRCA2 gene analysis in FFPE tumor samples. *BMC Cancer*. **19**, 396 (2019).
30. Ward, D. G. et al. Multiplex PCR and next generation sequencing for the non-invasive detection of bladder cancer. *PLoS ONE*. **11**, e0149756 (2016).
31. Gautam, R. et al. One-step multiplex real-time RT-PCR assay for detecting and genotyping wild-type group a rotavirus strains and vaccine strains (rotarix® and RotaTeq®) in stool samples. *PeerJ* **4**, e1560 (2016).
32. Liu, Z. et al. New multiplex real-time PCR approach to detect gene mutations for spinal muscular atrophy. *BMC Neurol.* **16**, 141 (2016).
33. Saito, T. et al. Rapid and sensitive multiplex single-tube nested PCR for the identification of five human plasmodium species. *Parasitol. Int.* **67**, 277–283 (2018).

Author contributions

J.Z. and Z.W. designed and supervised the study. Z.W., J.Z., S.S., W.W., and Y.Z. participated in sample collection, investigations, and data analysis. Z.W., J.Z., S.S., W.W., and Y.Z. led the data curation and processing. Z.W. led the data analysis and visualization. Z.W., J.Z., S.S., W.W., and Y.Z. led the results interpretation and writing of the manuscript. All authors contributed to the editing of the manuscript and approved the final version.

Funding

This work was supported by the China Agriculture Research System of Specialty Freshwater Fish (CARS-46) and the Yarlung Zangbo River Resources Survey.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-45045-8.

Correspondence and requests for materials should be addressed to Z.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.