



Development of single nucleotide polymorphisms markers associated with disease resistance towards *Pseudomonas koreensis* in empurau (*Tor tambroides*) using T-Plex ARMS PCR assay

Melinda Mei Lin Lau¹ · Hung Hui Chung¹ · Cindy Jia Yung Kho² · Han Ming Gan³ · Azham Zulkharnain⁴

Received: 27 February 2025 / Accepted: 21 May 2025
© The Author(s) 2025

Abstract

Empurau is a highly sought freshwater fish with high market value due to its unique flesh taste. However, the main challenge of the industry remains the infectious disease outbreak, with *Pseudomonas* species as one of the most threatening fish pathogens while residing in a wide range of environments. Antibiotic use is a common solution to disease outbreaks and leads to increased antimicrobial resistance. The development of disease-resistant broodstock through the identification of SNP markers has emerged as a promising strategy. In this study, a total of 1,048,576 SNP markers were identified via whole genome pooled sequencing on samples treated with LD₅₀, forming resistant and susceptible groups. Multiple tests (pairwise F_{ST} test, CMH test, and FE test) and visualization were conducted to screen and select candidate SNP markers for further validation using T-plex ARMS real-time PCR assay. The genotyping results on the selected candidate markers were confirmed using Sanger sequencing. Statistical analysis was performed to validate the significance of the candidate markers. The successful validation of 19,564 G/C SNP markers is hoped to contribute to research focusing on disease-resistant association SNPs, enabling the application of genomic information to enhance artificial selection strategies in *Tor tambroides* (empurau) breeding.

Keywords *Tor tambroides* · *Pseudomonas koreensis* · SNPs · Disease resistance · T-plex ARMS PCR assay

Introduction

Aquaculture is expanding faster as one of the rapidly growing food production sectors worldwide, driven by the increasing demand for fish products that provide proteins and essential micronutrients for balanced nutrition and good health (FAO 2016; DOF 2019; Jumatli and Ismail 2021). Fish production had achieved up to 17.3% of global population

Handling Editor: Amany Abbass

Extended author information available on the last page of the article

animals' protein intake and 6.8% of all proteins taken in year 2017, further emphasizing the importance of fish as the prime protein source. Furthermore, with the progressive improvement of the aquaculture industry and stagnant capture fisheries, world fish production is expected to increase up to 187 million metric tons (MT) to cope with the increasing world population (World Bank 2013). Malaysia's Fisheries Sector, including capture fisheries, inland fisheries, and aquaculture, produced up to 5.7 million tonnes with an estimated value of RM4 billion, with increments of 37.5% and 17.2% in quantity and value, respectively, as compared to year 2021 (DOF 2022).

Tor tambroides, or locally known as "empurau," is one of the prized species within the *Cyprinidae* family, valued not only for human consumption but also for its ornamental and sporting values. In addition to the excessive fishing pressure driven by its unique flesh taste, the species is also threatened by artificial propagation activities undertaken for both aquaculture production and conservation purposes (Ingram et al. 2005; Kottelat et al. 2018; Lau et al. 2022). In spite of the challenges, its high market price allows it to be one of the most prominent species in the aquaculture industry (Ingram et al. 2005). Nevertheless, its susceptibility to disease outbreaks in captivity due to overpopulation further encourages pathogen spreading (Magnadóttir 2010). There have been huge cases of viral and bacterial infections reported within local fish farms, further causing huge economic losses to the industry (Chiew et al. 2019).

Bacterial pathogens are the primary culprits behind most fish farming illnesses (Irsthath et al. 2023). Diverse strains of *Pseudomonas* species pose a significant threat to fish, leading to considerable losses in the aquaculture sector (Pękala-Safińska 2018). Specifically, *Pseudomonas koreensis* has been identified as a key contributor to the mortality of diseased *Tor tambroides*. The symptoms displayed were exophthalmia, abdominal swellings, loss of scales, and hemorrhaging found around the fins and gills (Lau et al. 2022, 2024). Besides, back in the year 2014, *P. koreensis* was documented as the causative agent for eye lesions in gold mahseer (*Tor putitora*) in India (Shahi and Malik 2014). Despite its limited disease occurrence in fish, public health focuses much of its attention toward *Pseudomonas* infection as it causes serious foodborne infection due to raw fish and byproduct consumption (Zilberberg and Shorr 2012). These findings are reinforced by reported cases of *Pseudomonas aeruginosa* isolated from fish, which have been linked to instances of pneumonia in hospital settings (Novotny et al. 2004) and the combination of *P. koreensis* and *Aspergillus fumigatus* in causing mixed infectious keratitis in humans (Khoo et al. 2021).

Upon disease outbreaks, the fastest and immediate solution is the usage of antibiotics. However, the extensive misuse of antibiotics can lead to immunosuppression, generation of resistant pathogen strains, environmental pollution, and even high risk to human health due to tissue deposition (Uddin et al. 2021; Rohani et al. 2022). While improving the water environment can help manage survival rates to some extent, strengthening disease resistance is recommended as a more effective and sustainable strategy for disease control in the aquaculture industry (Irsthath et al. 2023). Antibiotics, heralded as one of the most impactful medical breakthroughs of the twentieth century, have unquestionably played a vital role in combating bacterial infections, ultimately saving millions of lives (Ribeiro da Cunha et al. 2019). However, the occurrence of antimicrobial resistance (AMR) is still an unavoidable evolutionary result. Antibiotic wastes were released into the environment, further encouraging the formation of antibiotic reservoirs and antibiotic resistance genes along with their resistance genes (Manyi-Loh et al. 2018). Consequently, human society has to face the drawbacks as diseases including pneumonia, septicemia, tuberculosis, salmonellosis, and gonorrhea are becoming difficult to cure due to AMR (Bondad-Reantaso et al. 2023). In the year 2019, the World Health Organization (WHO) reported 700,000 deaths

due to AMR, and it is predicted the figure will be on the rise to reach 20 million by the year 2050 (Watkins et al. 2016).

The success and sustainability of the aquaculture industry hinge heavily on effective disease control measures. Developing disease-resistant broodstock has emerged as a promising strategy to address antimicrobial resistance (AMR) concerns in aquaculture (Moss et al. 2012). Growth in genomic technologies has enabled molecular markers such as single nucleotide polymorphism (SNP) to be readily available. It is cheaper to run in a dense panel as compared to microsatellites, thus allowing more individuals to be genotyped at a greater genome coverage level (Vignal et al. 2002). SNPs that are significantly associated with a certain trait can be utilized further for trait improvement through selective breeding. It has been applied in aquatic species to facilitate selective breeding and increase the rate of gene discovery in relation to economic traits, for instance growth (Lv et al. 2015); disease resistance (Yue et al. 2012; Zhang et al. 2019; Zhao et al. 2021); and body conformation (Geng et al. 2017). As the number of SNPs associated with the target traits continues to grow, a reliable approach for easy detection of these markers, not only in laboratory settings but also in practical aquaculture environments, remains crucial. In this study, pooled-sequencing (Pool-Seq) serves as a more cost-effective alternative for SNP identification. Several studies have implemented this approach to quantify allele frequency differences between populations to infer about the selection process (Boitard et al. 2012; Fischer et al. 2013; Bajpai et al. 2022). The T-plex real-time PCR assay utilizes a combination of techniques, including tetra-primer ARMS PCR assay, SYBR Green I-based real-time PCR, and melting point analysis. This approach incorporates a primer design strategy to specifically detect the SNP (single nucleotide polymorphism) of interest (Baris et al. 2013). It is able to distinguish the amplified allele-specific amplicons in a single tube via melting temperature. Furthermore, pairwise F_{ST} , Fisher exact test, and Cochran-Mantel Haenszel (CMH) test were implemented for SNP marker candidate selection which reflects allele frequency differences and their respective genetic differentiation between both resistant and susceptible groups.

Considering the significant economic and cultural value of *Tor tambroides*, any disease occurrence could potentially lead to severe repercussions. Therefore, in this study, we aimed to detect and identify candidate disease-resistant SNPs towards *Pseudomonas korensis* by performing whole genome Pool-Seq on both groups of fish. These findings are hoped to contribute to further studies considering disease-resistant association SNPs and can be used in breeding programs to integrate genomic information into artificial selection decisions in emperau breeding.

Materials and methods

Bacterial challenge and sample preparation

For the bacterial challenge, 60 juvenile *Tor tambroides* were obtained from a local aquaculture farm (GPS coordinates: 1° 32' 53.647''N, 110° 32' 53.233''E) where the bacterial strain was originally isolated from. The bacterial strain was previously isolated from a diseased adult *T. tambroides* (intestine and liver) sampled from the local aquaculture farm on *Aeromonas* isolation agar (Sigma Aldrich; USA) supplemented with ampicillin (0.5 mg/mL) (Lau et al. 2024). The juvenile fishes were randomly distributed into six tanks equipped with filters and aeration. Treated tap water was used, and twice water changes

(40% of the total water volume) were done per week. The fish were fed with commercial pellets once daily in the morning.

The bacteria strain used was *Pseudomonas koreensis* CM-01 isolated from diseased *Tor tambroides* (Kho et al. 2023; Lau et al. 2024). Five of the six groups were introduced to *P. koreensis* at 5×10^7 CFU/mL (LD_{50}) while the remaining group acted as a control group (Lau et al. 2024). After the introduction of *P. koreensis* and $1 \times$ PBS (control group) into the fish via peritoneal injection, all fish were raised for 7 days at 25 °C for observation. All the experimentally infected fish were observed at hourly intervals for the first 12 h and at 6-h intervals for the subsequent observation period. Moribund or dead fish observed during the post-infection challenge were categorized as the susceptible group (SG) while those fish that survived the challenge were considered the resistant group (RG). Twenty samples from SG and 20 samples from RG were collected randomly, and the genomic DNA was extracted from muscle tissue and later stored at -80 °C. The genomic DNA extraction was done in replicates via the modified CTAB-based extraction method. All experiments comply with ARRIVE guidelines and were carried out in compliance with the guidelines and permission approved by the Animal Ethics Committee of Universiti Malaysia Sarawak (UNIMAS/TNC(PI)–04.01/06–09(17)).

Sequencing and read cleaning

In this study, a pooled sequencing approach was implemented. A total of 40 individuals, consisting of 20 resistant and 20 susceptible individuals, were selected for the study. The susceptible group primarily included individuals displaying severe clinical symptoms such as pop eyes, inflamed vents, and petechial hemorrhages. Extracted DNA from both resistant and susceptible groups was pooled to a final concentration of 40 $\mu\text{g}/\mu\text{L}$, with each DNA sample contributing 2 $\mu\text{g}/\mu\text{L}$ (Bajpai et al. 2022; Baltrusis et al. 2022; Rellstab et al. 2013). Approximately 1 μg of gDNA was sheared to 350 bp using a Bioruptor and directly used for PCR-free library preparation using the NEB Ultra Illumina library preparation kit (NEB, Ipswich, MA). The library was quantified with a Qubit (Invitrogen) and sequenced on a NovaSEQ6000 (Illumina, San Diego, CA) with 2×150 bp run configuration. The sequences were trimmed off the adapter sequences by fastp version 0.23.0 before proceeding with SNP calling (Chen et al. 2018).

Reads mapping, SNP calling, and identification

Trimmed paired-end reads of both resistant and susceptible reads were mapped to the *Tor tambroides* reference genome (GCA_021397915.1) (Lau et al. 2022). The reference genome was prepared prior to sequence alignment using the Burrows-Wheeler aligner (BWA) *index* command (Li and Durbin 2010). The reads were mapped using the BWA *mem* command. Samtools version 1.13 (Danecek et al. 2021) was then used to sort the BAM file and subsequently create an mpileup file containing outputs from both sample pools. Subsequently, the mpileup file was subjected to PoPoolation2 analysis (Kofler et al. 2011). A sync file was created using the PoPoolation2 script mpileup2 sync.jar by keeping the base quality at 20. Next, SNPs' allelic frequency (script: snp-frequency-diff.pl) was used by maintaining a minimum allele count of 6, and minimum and maximum coverages of 50 and 200, respectively. To investigate the presence of any consistent allele frequency changes belonging to the two sample groups, Cochran-Mantel–Haenszel (CMH) tests were conducted (script: cmh-test.pl). Nevertheless, due to the pooled samples in both sample

groups, they do not represent true, distinct pairs and were compared in an arbitrary fashion. Thus, the significance of per nucleotide allele frequency differences between the two groups was determined using Fisher's exact (FE) test (script: fisher-test.pl). Pairwise fixation index (F_{ST}) values were determined throughout the genome (script: fst-sliding.pl). The filtered list of SNP markers was visualized using the Integrative Genomics Viewer (IGV) to further refine the selection.

T-Plex real-time PCR assay

A total of five potential SNPs were targeted in this assay. The choices of the SNPs were identified through the comparison of the SNPs between the resistant group, susceptible group, and reference genome (GCA_021397915.1) (Lau et al. 2022). The reference genome serves as a control in the selection process, enabling more precise and reliable identification of SNPs. Potential SNP markers were selected based on unique nucleotide patterns that differentiated them from both the susceptible group and the reference genome. The location of the selected five SNP markers was predicted using web-based Augustus (<https://bioinf.uni-greifswald.de/augustus/submission.php>) (Stanke and Morgenstern 2005). Next, nucleotide and predicted protein sequences where each SNP marker is located were functionally annotated using NCBI conserved domain (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) (Wang et al. 2023), Interproscan (<https://www.ebi.ac.uk/interpro/search/sequence/>) (Jones et al. 2014), NCBI BLAST (Wheeler et al. 2007), and eggNOG-mapper (<http://eggnog-mapper.embl.de/>) (Cantalapiedra et al. 2021). This approach aimed to elucidate the underlying biological roles of the selected SNP markers involved.

For SNP genotyping, T-plex ARMS real-time PCR assay was used (Baris et al. 2013). This assay is a combination of tetra-primer amplification-refraction mutation assay (ARMS) PCR assay, SYBR Green I-based real-time PCR, and melting-point analysis with primer design strategy to detect the SNP of interest (Baris et al. 2013) (Fig. 1A, B).

Primer design

The T-Plex real-time PCR assay was used to genotype the 6151 C/T, 112,388 G/A, 59747G/A, 27857 A/T, and 19564G/C. Primers were designed based on the sequences available on the public databases (SRX21910145 and SRX21910146) using the web-based software (<http://primer1.soton.ac.uk/primer1.html>) (Ye et al. 2001; Collins and Ke 2012). Next, the specificity of the primers was checked using NCBI BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). The mismatch was introduced at the third position of the 3' end of both allele-specific primers to increase the specificity of the reaction, which is automatically introduced by the software. The hypothetical T_m values of the primers were calculated using IDT OligoAnalyzer (<https://sg.idtdna.com/calc/analyzer>). The primers used in this study are listed in Table 1.

SNP validation

Real-time PCR was conducted using Thunderbird Next SYBR qPCR Mix (Toyobo; Osaka, Japan) and QuantStudio 5 (ThermoFisher; Waltham, Massachusetts, USA). The total volume of a single reaction mixture is 20 μ L, containing 10 μ L of qPCR mix, 1 μ L of each primer per reaction (500 nM), and 1 μ L of genomic DNA dilution (10 ng/ μ L)

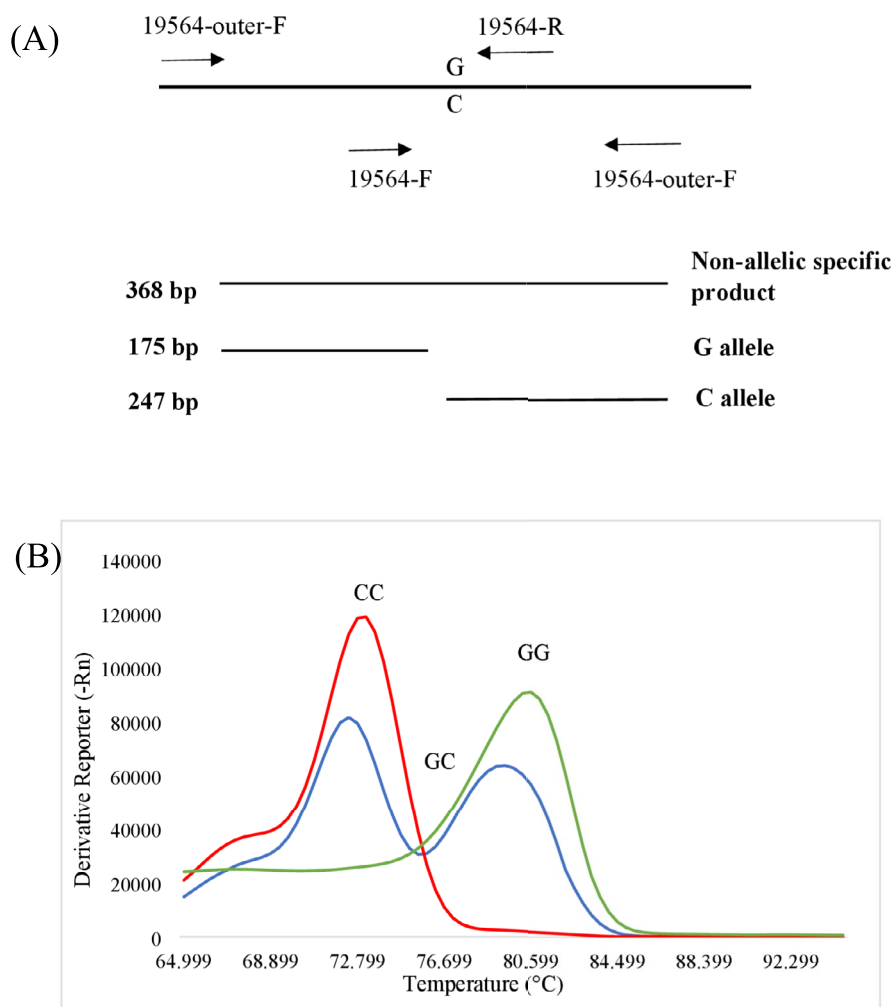


Fig. 1 T-Plex real-time PCR assay. **A** It shows a schematic diagram of tetra-primer ARMS PCR for 19,564 G/C genotyping in this study as an example. During the PCR amplification, a 368-bp amplicon will be amplified by the non-allelic outer primer, and the allele-specific amplification produced 175- and 247- PCR products specific for the allele G and allele C, respectively. **B** T-Plex real-time PCR assay coupled with melting curve analysis for the 19,564 G/C SNP marker genotyping using SG and RG samples. The genotyping was performed based on amplicons produced at specific T_m , with the red line representing CC genotype, the green line representing GG genotype, and the blue line representing GC genotype

and ultrapure water. The qPCR was conducted as follows: an initial denaturation (95 °C for 7 min), followed by amplification and quantification steps repeated for 45 cycles (95 °C for 10 s, 60 °C for 10 s, 72 °C for 20 s), with a single fluorescence measurement at the end of the final elongation step. Next, the melting-curve analysis was conducted by reducing the temperature to 65 °C and increasing the temperature by 0.2 °C/s to 95

Table 1 Primer sequences for potential SNP validation used in this study, designed using a web-based software

SNP (Contig)	Sequences	Expected size (bp)	SNP location	SNP type
6151 C/T (WUC47414__np12)	F	413	6151	C/T
	R	-		
	C	246		
	T	184		
	F	364	112,388	
112,388 G/A (WSC241__np12)	R	-		G/A
	A	246		
	G	156		
	F	387	59,747	
	R	-		
59,747 G/A (P_RNA_scaffold_1984__np12)	R	247		G/A
	G	181		
	A	358	27,857	
	F	-		
	R	247		
27,857 (WSC6179__np12)	A	161		A/T
	F	368	19,564	
	R	-		
	G	175		
	C	247		
19,564 (WSC72258__np12)	A	247		G/C
	T	161		
	F	368	19,564	
	R	-		
	G	175		

For the allele-specific primer pairs, the nucleotide in italics is the mismatch introduced, while the nucleotide in bold targets the SNP for both resistant and susceptible populations

°C while continuously measuring the change in fluorescence. The reaction was terminated by cooling to 40 °C. Genotypes were identified via the melting curves and the melting-peaks display on the QuantStudio™ Design & Analysis Software version 1.5.3. For verification of the genotypes identification results of the qPCR assay, six samples (three samples from RG and three samples from SG) were chosen to be amplified using non-allelic specific primers. The PCR products were sequenced, and the chromatogram of the sequencing results was visualized and analyzed using QSV analyzer (Carr et al. 2009).

Results

General overview and SNP identification

A total of 50.87 Gb and 50.48 Gb of clean data were obtained from the RG and SG sequencing libraries, respectively (Table 2). Both sample groups had shown high mapping rates of 93.66% and 93.38% with 337.76 M and 314.71 M reads, respectively. The trimmed sequencing data from both RG and SG were aligned to the reference genome before being subjected to sync file generation before PoPoolation2 analysis. A total of 1,048,576 SNPs had been successfully called from the sync file generated from both groups.

Among them, 629,670 SNPs were putative transitions (include four classes: A/G, G/A, C/T, and T/C) and 390,111 SNPs were putative transversions (include eight classes: A/C, C/A, A/T, T/A, G/C, C/G, G/T, and T/G) (Fig. 2). The distribution of transition SNPs was higher than that of transversion SNPs by twofold. G/A and C/T were the most commonly observed transition SNPs, recorded at 173,343 and 173,903, respectively. As for the transversion SNPs, there were 64,512 A/T and 63,510 T/A SNPs observed.

Pairwise F_{ST} test, Fisher's exact (FE) test, and Cochran-Mantel-Haenszel (CMH) test were performed for selecting SNPs prior to T-plex real-time PCR assay. The SNPs were filtered from the top 1% F_{ST} values and p-value from the FE test along with p-value < 0.001 in the CMH test. Nucleotide distribution among both groups was compared for the subsequent validation step (Table 3). Table 3 showed the results of this study: the selected SNP markers along with their respective contig and parameters, including the CMH test, pairwise F_{ST} value, and FE test. All selected SNPs have p-value less than 1.0×10^{-13} and 1.0×10^{-9} for CMH test and FE test, respectively. Candidate SNP markers were further refined using Integrative Genomics Viewer (IGV), selecting those with unique coverage and distribution patterns in the resistant group compared to the susceptible and control groups (Figs. S1, S2, S3, S4, S5).

Table 2 Summary statistics for the resistant group and the susceptible group from pooled sequencing

Parameter	Resistant group (RG)	Susceptible group (SG)
Total length (bp)	50.87 Gb	50.48 Gb
Map read rate (%)	93.66	93.38
Reads (M)	337.76	314.71
Sequencing depth	42.39	42.07
Total SNPs (sync file)	1,048,576	

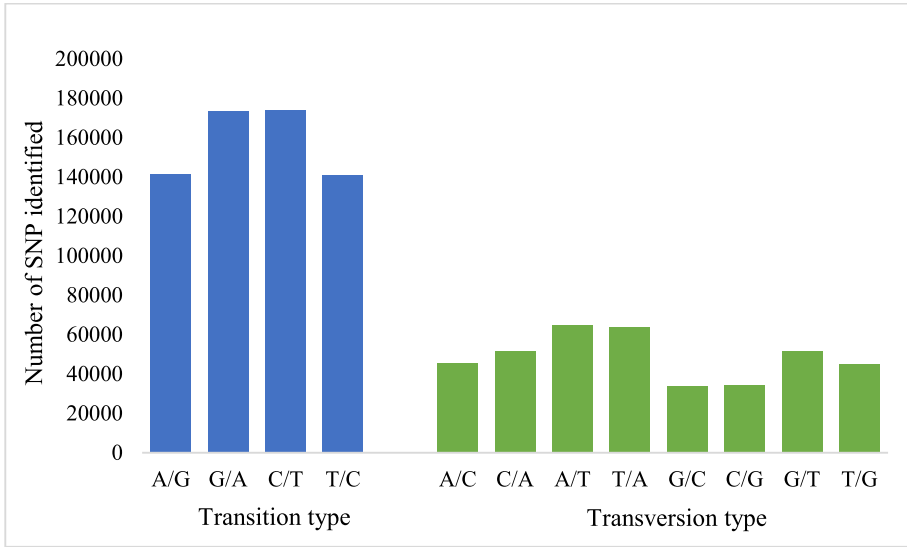


Fig. 2 Classification of SNP identification into transition and transversion types identified within the genome of *T. tambroides*

Table 3 SNP selections including CMH test, pairwise F_{ST} , and FE test were performed for further validation using T-Plex real-time PCR assay

SNP	Cochran-Mantel-Haenszel test			Pairwise F_{ST}	Fisher's exact test
	RG (A:T:C:G:N)	SG (A:T:C:G:N)	<i>p</i> -value		
6151 C/T	0:0:99:0:0:0	0:26:42:0:0:0	1.58×10^{-20}	0.2363	2.22×10^{-12}
112,388 G/A	0:0:0:70:0:0	27:0:0:50:0:0	3.51×10^{-15}	0.2136	1.92×10^{-9}
59,747 G/A	0:0:0:51:0:0	28:0:0:32:0:0	6.80×10^{-14}	0.3064	8.03×10^{-10}
27,857 A/T	0:113:0:0:0:0	23:76:0:0:0:1	6.38×10^{-14}	0.1314	5.33×10^{-9}
19,654 G/C	0:0:0:114:0:0	0:0:27:97:0:0	2.42×10^{-13}	0.1228	5.31×10^{-9}

N refers to any base A, T, C, or G

Functional annotation

All SNP markers were found to fall within gene and protein categories of KEGG Biomolecular Relations in Information Transmission and Expression (BRITE) and all were correlated with metabolism (Table 4). Specifically, 6151 C/T SNP was responsible for transcription machinery, DNA repair and recombination proteins, enzymes, ubiquitin system, and RNA polymerase in *T. tambroides*. Furthermore, it was found to be associated with posttranslational modification, protein turnover, and chaperones, which fall under the category cellular processes and signalling in COG databases. It possesses DUF5401 superfamily and B-box-type zinc finger as domain as well. In NCBI, it was found to have 90.6% similarity to E3 ubiquitin protein ligase TRIM35-like.

The second SNP marker, 112,388 G/A, falls within the coding region which encodes for adrenoceptor alpha 2 C (97.23%). NCBI conserved motif and Interproscan have reported

Table 4 Functional annotations of the five selected SNP markers

SNP (location)	COG	KEGG BRTE (gene and proteins)	KEGG pathway	KEGG ortholog
6151 C/T (intron within gene 1)	Posttranslational modification, protein turnover, chaperones	Enzymes Transcription machinery DNA repair and recombination proteins Ubiquitin system RNA polymerase	RNA Polymerase Nucleotide excision repair Huntington disease	RPB1 TRIM35 TRIM39
112,388 G/A (coding region at gene 7)	Transcription Replication, recombination and repair Signal transduction mechanism Function unknown	Peptidase and inhibitors Enzymes Transcription factors G protein-coupled receptors	cGMP-PKG signalling pathway Neuroactive ligand-receptor interaction Type I diabetes mellitus	GPR26 CPN1 NKX5 ADRA2 A, ADRA2B, ADRA2 C CPE, CPD, CPZ AEPB1
59,747 G/A (intron within gene 1)	Cytoskeleton	Enzymes Exosomes Cytoskeleton proteins	cGMP-PKG signalling pathway focal adhesion Salmonella infection Proteoglycans in cancer Cell adhesion molecules	FLNA PARP7S NRCAM
27,857 A/T (intron within gene 2)	Replication, recombination and repair Signal transduction mechanism	Membrane trafficking Cell adhesion molecules Glycosylphosphatidylinositol (GPI)-anchored proteins Enzymes	-	CNTN2, CNTN3, CNTN4, CNTN5 USP22_27_51
19,654 G/C (intergenic sequences)	Replication, recombination and repair	Enzymes	-	USP22_27_51

Table 4 (continued)

SNP (location)	COG	KEGG BRITE (gene and proteins)	KEGG pathway	KEGG ortholog
	Posttranslational modification, protein turnover, chaperones	Peptidase and inhibitors		USP35_38
	Function unknown	Mitochondrial biogenesis		USP_26_37
		Transcription machinery		PI F1
		Chromosome and associated proteins		
		Ubiquitin system		
		DNA replication proteins		

The location of SNP markers was predicted using Augustus, while the annotations to COG and KEGG databases were performed using eggNOG-mapper, using predicted protein sequences

RPBI, DNA-directed RNA polymerase II subunit *RPBI*; *TRIM35*, tripartite motif-containing protein 35; *TRIM39*, tripartite motif-containing protein 39; *GPR26*, G protein-coupled receptor 26; *CPNI*, carboxypeptidase N catalytic subunit; *MKX5*, homeobox protein *Nkx-5*; *ADRA2 A*, adrenergic receptor alpha-2 A; *ADRA2B*, adrenergic receptor alpha-2B; *ADRA2 C*, adrenergic receptor alpha-2 C; *CPE*, carboxypeptidase E; *CPD*, carboxypeptidase D; *CPE*, carboxypeptidase Z; *FLMA*, filamin A; *PARP75*, poly [ADP-ribose] polymerase; *NRCAM*, neuronal cell adhesion molecule; *CNTN2*, contactin 2; *CNTN3*, contactin 3; *CNTN4*, contactin 4; *CNTN5*, contactin 5; *USP22_27_51*, ubiquitin carboxyl-terminal hydrolase 22/27/51; *USP35_38*, ubiquitin carboxyl-terminal hydrolase 35/38; *USP26_29_37*, ubiquitin carboxyl-terminal hydrolase 26/29/37; *PIF1*, ATP-dependent DNA helicase *PIF1*

on the presence of the alpha2 C subtype adrenoreceptor. It is associated with cGMP-PKG signalling and neuroactive ligand-receptor interaction pathway in KEGG (Table 4). As for COG databases, their categorization mainly falls within information storage and processing, including transcription, signal transduction mechanism, replication, recombination, and repair.

Subsequently, the 59,747 G/A SNP marker is an intron within a predicted gene, filamin-C-like gene (NCBI similarity 96.69%). As reported from InterProScan and NCBI conserved domain, it had filamin/ABP280 repeat containing protein and filamin type immunoglobulin domain. It was found to be interacting with four KEGG pathways: MAPK signalling pathway, focal adhesion, *Salmonella* infection, and proteoglycans in cancer. This can be further supported by its classification, which mainly focuses on signalling and cellular processes (enzymes, exosomes and cytoskeleton proteins) by both KEGG and COG databases.

The fourth SNP marker, 27,857 A/T, falls as an intron, which was characterized as contactin-5-like by NCBI (97.55%). Both InterProScan and NCBI conserved motif reported fibronectin type III and immunoglobulin-like fold as its domain, which was categorized as cell adhesion molecules (Table 4). The 27,857 A/T SNP was found to be associated with glycosylphosphatidylinositol (GPI)-anchored proteins, cell adhesion molecules, and membrane trafficking in KEGG BRITE. It was classified as information storage and processing, which is responsible for replication, recombination, and repair, as well as signal transduction mechanism, as reported in the COG database.

Lastly, 19,564 G/C SNP falls outside the transcript region as intergenic sequences and its contig functions as zinc finger and SCAN domain-containing protein 29-like in NCBI (94.02%) and its function may be regulated by Myb/SANT-like DNA binding domain. It was not found in any KEGG pathways, but it did find an association with ubiquitin carboxyl-terminal hydrolase and ATP-dependent DNA helicase (Table 4). Within KEGG BRITE, it was mainly classified as metabolism and genetic information processing, its functions focusing on enzymes, peptidase and inhibitors, mitochondrial biogenesis, transcriptional machinery, chromosome and associated proteins, ubiquitin system, and DNA replication proteins. COG databases reported 19,564 G/C SNP to fall under posttranslational modification, recombination, and repair as well as the replication, recombination, and repair category.

SNP validation via T-Plex real-time PCR assay

Twenty samples in SG and 20 samples in RG were used to confirm the genotype of the candidate SNP markers. A total of five SNP markers were selected to verify and four SNPs were successfully verified except 59,747 G/A SNP due to its unsuccessful PCR amplification (Figures S6, S7, S8, S9, S10). The significant different alleles from the SNP markers were detected between both RG and SG using X^2 test with one SNP marker successfully shown significant differences between both groups (Table 5). Out of the verified SNP markers, only 6151 C/T SNP showed one genotype for both RG and SG, which is the CC genotype. Furthermore, the same genotype distribution was observed for both 112,388 G/A and 27,857 A/T SNP, with most of the individuals showing similar homozygous genotype in both groups, eventually bringing to insignificant statistical differences in the particular SNP for both populations. 19,564 G/C SNP shows greater GG genotype distributions in RG while in SG, greater CC genotype distributions were observed. The T_m graphs of each SNP marker for both RG and SG are summarized in Figs. S6, S7, S8, S9, S10.

Table 5 Distribution of the five markers in the susceptible and resistant groups within *T. tambroides*

SNP	Genotype	RG (n)	SG (n)	X ² (p)
6151 C/T	CC	20	20	-
	CT	0	0	
112,388 G/A	GG	20	17	3.24 (0.072)
	AG	0	3	
27,857 A/T	TT	20	17	3.24 (0.072)
	AT	0	3	
19,564 G/C	GG	16	4	14.01 (0.000907)
	CC	1	13	
	GC	3	3	

For further verification, Sanger sequencing results of four SNP markers (6151 C/T, 112,388 G/A, 27,857 A/T, and 19,564 G/C) had portrayed that the SNP genotypes are consistent with the grouping results of T-plex real-time PCR analysis (Fig. 3) with the red-colored line graph representing RG and the blue-colored line graph representing SG. For the 6151 C/T SNP, selected RG samples (R16, R17, R20) and SG samples (S7, S10, S12) had shown T_m of around 76 °C (Fig. 3B) which can be further supported by the multiple alignment of Sanger sequencing results showing that the T_m belongs to the C allele (Fig. 3A). As for the second SNP 112388 G/A, all RG samples (R1, R16, R20) and samples from SG (S14, S17) had shown a single T_m peak at 87 °C. However, sample S20 from SG showed a double T_m peak at 87 °C and 77 °C for allele G and allele A, respectively (Fig. 3C, D).

As for 27,857 A/T SNP, all selected samples from RG (R1, R12, R15) and SG (S1, S2, S12) had shown a single T_m peak at 74 °C, which is supported by Sanger sequencing results (Fig. 3E, F). Selected RG samples (R2, R8, R17) had shown double T_m peaks at 79 °C and 72 °C. As for the SG, three samples had shown different patterns of T_m peaks, with sample S1 showing double peaks at 79 °C and 72 °C, and S2 and S20 showing single peaks at different T_m of 79 °C and 72 °C, respectively (Fig. 3G, H).

With the successful genotyping of 19,564 G/C SNP, Table 6 had summarized the genotype verification using Sanger sequencing in comparison with the melt curve from real-time PCR. For RG samples, selected samples R2, R8, and R17 all show GC genotype due to the double nucleotide peak detected from the Sanger sequencing chromatogram in QSV analyzer. Besides, it is found consistent with the double T_m peak of real-time PCR at 79 °C and 72 °C. As for the SG, sample S1 showed a double peak in both the Sanger sequencing chromatogram and melt curve, which confirmed its GC genotype. Furthermore, samples S2 and S20 portrayed a single peak with different nucleotide and T_m , with allele G at 79 °C and allele C at 72 °C, respectively. Thus, the genotype for 19,564 G/C SNP can be deduced via the T_m shown by each sample and is applied for all the SNP markers (Fig. S5) (Table 6).

Discussion

T-Plex ARMS PCR assay and pooled sequencing

In this study, we had pooled and sequenced the whole genome of two pools of DNA, resistant group and susceptible group for the identification of disease resistant SNP

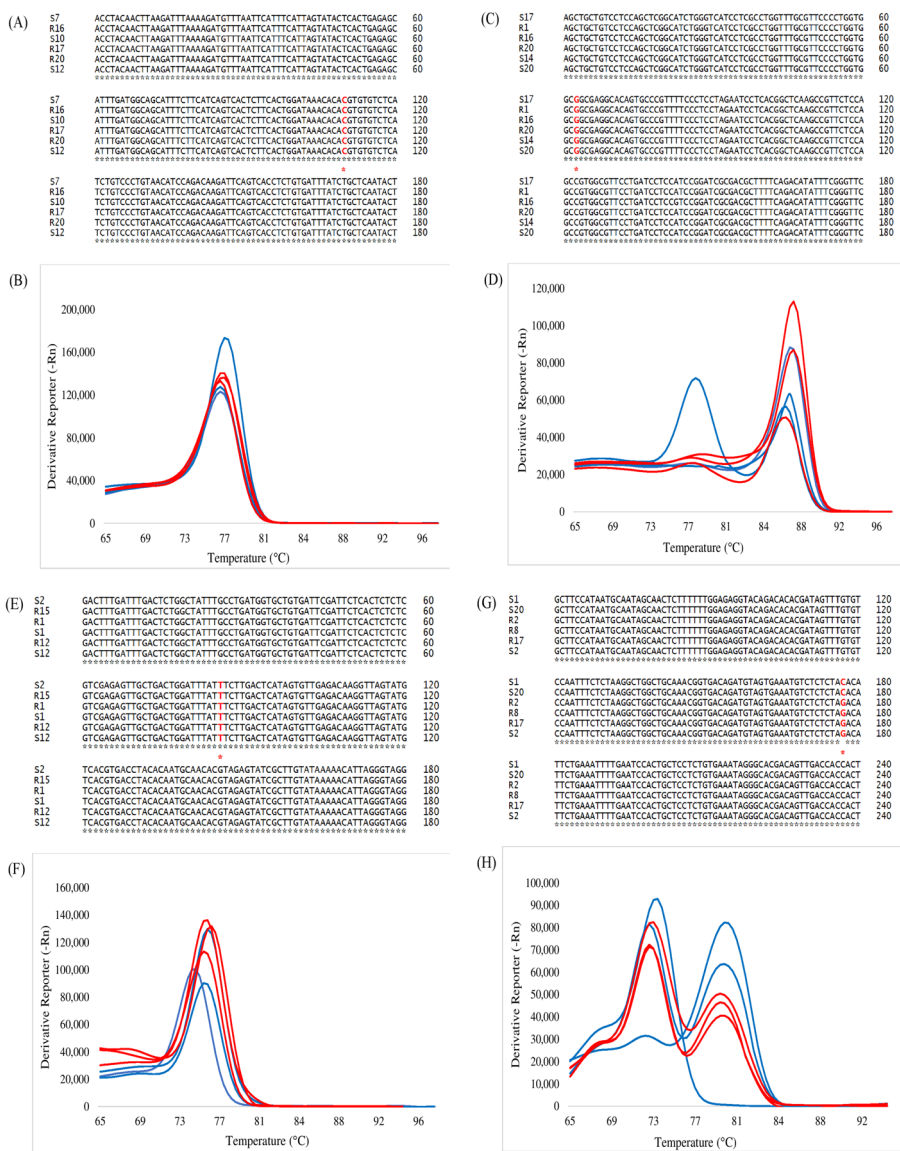
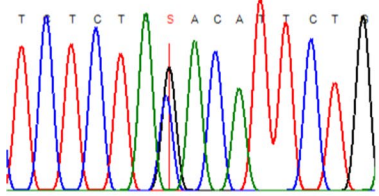
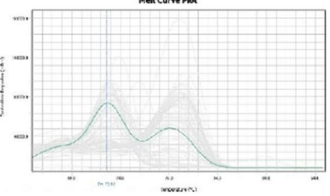
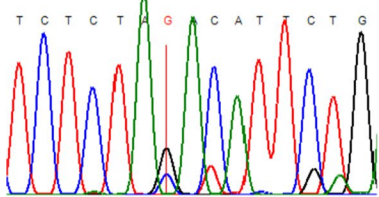
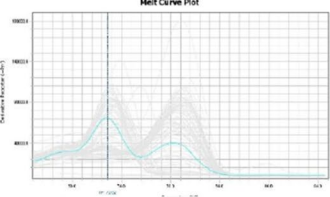
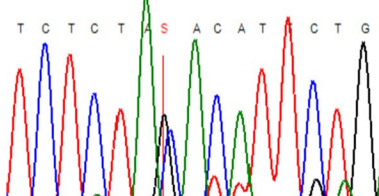
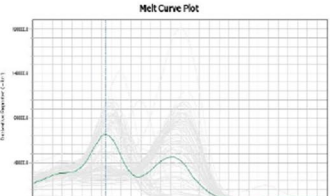
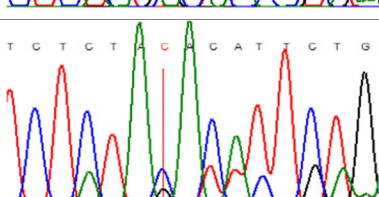
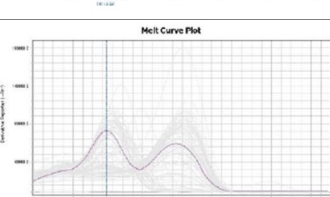
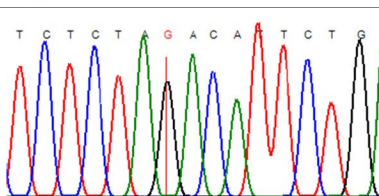
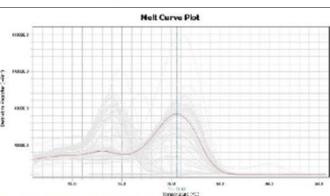
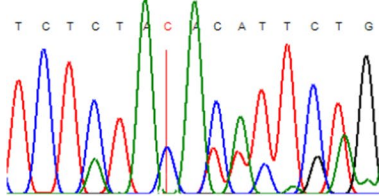
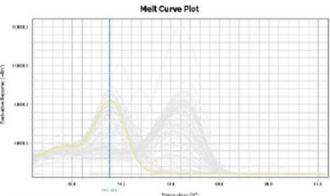


Fig. 3 The results of Sanger sequencing and T-plex real-time PCR analysis on 6151 C/T, 112,388 G/A, 27,857 A/T, and 19,564 G/C SNPs. **A** Multiple sequence alignment of PCR products of 6151 C/T SNP for six samples. **B** Discrimination of 6151 C/T SNP marker among six corresponding samples. **C** Multiple sequence alignment of PCR products of 112,388 G/A SNP for six samples. **D** Discrimination of 112,388 G/A SNP marker among six corresponding samples. **E** Multiple sequence alignment of PCR products of 27,857 A/T SNP for six samples. **F** Discrimination of 27,857 A/T SNP marker among six corresponding samples. **G** Multiple sequence alignment of PCR products of 19,564 G/C SNP for six samples. **H** Discrimination of 19,564 G/C SNP marker among six corresponding samples. A red star represents SNP. RG represents resistance group (red-colored line graph) while SG represents the susceptible group (blue-colored line graph)

Table 6 Genotype verification via Sanger sequencing while in comparison with melt curve from real-time PCR on 3 samples each from both RG and SG on 19,564 G/C SNP

Sample (Genotype)	Sanger Sequencing Chromatogram	Real-time PCR Melt Curve
R2 (GC)		
R8 (GC)		
R17 (GC)		
S1 (GC)		
S2 (GG)		
S20 (CC)		

markers towards *Pseudomonas koreensis*. A total of 50.87 Gb and 50.48 Gb clean data were obtained with mapping rates of 93.66% and 93.38% from RG and SG, respectively. A sync file was created from both groups before being subjected to SNP identification using PoPoolation2. A total of 1,048,576 SNPs were identified from the sync file. Pairwise F_{ST} , Fisher exact (FE) test, and Cochran-Mantel Haenszel (CMH) test were performed to select candidate SNP markers for further validation (Baltrušis et al. 2022; Bankers et al. 2017; Bajpai et al. 2022). Allele frequency differences were calculated with its degree of differences being measured by F_{ST} pairwise test, where a higher F_{ST} value indicates greater differentiation between both groups. Pairwise F_{ST} test was performed to identify genes with especially high levels of genetic differentiation between resistant and susceptible groups, which provide a set of potential SNP marker candidates for subsequent validation (Bankers et al. 2017). In addition, FE test aimed to focus on SNP markers exhibiting significant genetic differentiation between resistant and susceptible groups by comparing each allelic frequency (Bajpai et al. 2022) while CMH test was performed to determine the presence of any consistent allele changes (Baltrušis et al. 2022).

Pooled sequencing offers the advantage of predicting allele frequencies of SNP markers with reasonable accuracy, particularly at a substantially lower cost compared to individual genotyping. This cost-effectiveness stems from its ability to achieve higher total coverage, even though the accuracy may be relatively modest, especially for smaller or less accurate datasets (Schlötterer et al. 2014; Guirao-Rico and González 2021). The accuracy of allele frequency estimation via the Pearson coefficient had shown up to 0.95 when two pools of samples achieved an average of $40 \times$ sequencing depth (Dagnachew et al. 2022) which is also shown in this study with RG and SG recorded as 42.39 and 42.07, respectively (Table 2). Nevertheless, the effect of unequal distribution of individuals into the pool can be decreased by having accurate equimolar pooling of each genomic DNA and a balanced number of pooled samples (Gautier et al. 2013; Konczal et al. 2014). Pooled sequencing, as one of the cost-effective alternatives in SNP markers identification, had been reported across various species, including white shrimps (Zhang et al. 2019), geese (Ren et al. 2021), barber's pale worm (Baltrušis et al. 2022), freshwater snails (Bankers et al. 2017), rocket plant (Bajpai et al. 2022), and rainbow trout (Al-Tobasei et al. 2017).

In this study, T-Plex ARMS PCR assay was used to validate the selected SNP marker in association with disease resistance towards *Pseudomonas koreensis*. Tetra-primer ARMS PCR amplifies both the wild type and mutant alleles, along with a control fragment, within a single PCR tube. This method utilizes four primers to amplify a large, non-allelic-specific fragment containing the SNP sites, as well as allelic-specific amplicons representing each of the two allelic forms. The primer pairs amplifying the large non-allelic-specific fragment are known as outer primers, while the primer pairs containing a mismatch site (allelic-specific) at their 3' end are the inner primers. Together, the inner primers and the outer primers can simultaneously amplify both wild-type and mutant amplicons (Fig. 1A). Within a single reaction tube, it is expected to produce two or three PCR products via this tetra-primer ARMS PCR for genotyping both homozygous and heterozygous individuals, respectively. The accumulation of amplicons in the reaction can be tracked over time using the fluorescence dye SYBR Green and the LightCycler System. The amplicons produced by the allelic-specific inner primers typically exhibit distinct melting temperatures, which vary based on factors such as their GC content, length, and sequences (Fig. 1B). Thus, the genotyping can be performed

based on the T_m of the specific amplicons and the unique shape of the melting peaks (Baris et al. 2013).

Identification of disease-resistant SNPs against *Pseudomonas koreensis*

Out of the five selected SNP markers for further validation, all SNP markers were from the intron region except for 112,388 G/A and 19,564 G/C SNP which are positioned in the coding region and intergenic region, respectively. Studies had emphasized the ability of intron to increase gene expression independently of their role as transcription factor binding sites across numerous eukaryotic organisms. This phenomenon can be termed as “intron-mediated enhancement,” allowing intron to boost transcript level by influencing transcription rate, nuclear export, and transcript stability. In addition, introns have been found to enhance mRNA translation efficiency as well (Shaul 2017).

A study revealed that the majority of the disease-associated SNPs (daSNPs) of up to 93% were located outside the coding region, including introns, long terminal repeats (LTRs), and intergenic region (IGR) (Maurano et al. 2012; Chen and Tian 2016). This posed a challenge to the scientific community to interpret their involvement in diseases. Researchers had emphasized the common location of the non-coding region near the regulatory region, which further suggests their possible interference onto the host regulatory elements (Maurano et al. 2012). Chen and Tian (2016) had summarized that up to 60.5% of IGR daSNP-disease phenotype associations can be correlated through the functional relevance of the predicted genes to the corresponding disease genes.

DUF5401 superfamily and B-box-type zinc finger are the domains identified within the contig housing 6151 C/T SNP marker. Zinc finger family proteins form multiple finger-like protrusions for zinc binding and further interact with DNA, RNA, and protein. B-box-type zinc finger domain is one of the representatives containing one or more B-box proteins which bind to Zn ion to stabilize specialized tertiary structures (Klug and Schwabe 1995). Alpha 2C adrenoceptor was characterized within the contig where the 112,388 G/A SNP marker was chosen from. It belongs to the subcategory of G-protein-coupled receptors (GPCRs) which are responsible for the signal transduction across eukaryotic cell membranes and its vast range of signals includes light, hormone, olfactory, and gustatory stimuli as well as neurotransmitters (Lefkowitz 2004). In addition, the alpha 2C adrenoceptor is reported to associate with deficit hyperactivity disorder (AHDH) (Cho et al. 2008) and behavioral responses including psychiatric disorder and suicide completion (Rivero et al. 2016), as well as colon cancer (Njeim and Eid 2018).

The contig containing the 59,747 G/A SNP marker was characterized as part of a filamin-C-like (FLNC) gene with a filamin domain. FLNC serves as one of the filamin isoforms (A, B, C) which form cross-linkage with actin filaments and interact with various binding partners (Stossel et al. 2001). Besides maintaining the mechanical integrity of muscle cells, FLNC was involved in multiple cellular processes including mechanoprotection, intracellular signaling pathways, and actin remodeling as well as cell–cell and cell–matrix adhesion (Feng and Walsh 2004; Fujita et al. 2012). A past study had reported on its nonsense mutation which enlarged the heart due to the rupture of the myocardial wall and degeneration of skeletal muscle (Fujita et al. 2012). Fibronectin III and immunoglobulin-like fold (Ig) were characterized as the domain of contactin-5-like (CNTN5) gene found within the contig localized 27,857 A/T SNP marker. It has been found to be associated with various neuropsychiatric traits including autism spectrum disorder (ASD) (Fernandez et al. 2017), anorexia nervosa (Nakabayashi et al. 2009), substance abuse (Nikpay et al. 2012), and attention-deficit

hyperactivity disorder (Lionel et al. 2011). They are crucial in neuronal migration, synaptogenesis, regulation of neurite outgrowth, axon guidance, cell survival, neuron-glia interaction, and myelin formation (Oguro-Ando et al. 2017; Mohebiany et al. 2014).

Myb/SANT-like DNA binding domain (MSANTD) was found within the contig where the 19,564 G/C SNP marker is located. This domain in *Danio rerio* is orthologous to the human MSANTD, known for its role in positively regulating DNA-templated transcription and localized within nuclear bodies (Sayers et al. 2021). The SANT domain, a 50-amino-acid motif, acts as a nuclear receptor co-repressor structurally resembling the DNA-binding domain (DBD) of Myb proteins (Ogata et al. 1994; Aasland et al. 1996; Tahirov et al. 2001). It consists of switching-defective protein (Swi3), adaptor 2 (Ada2), nuclear receptor co-repressor (N-CoR), and transcription factor (TFIIIB) which acts as a unique histone-interaction module coupling histone binding to enzyme catalysis in chromatin remodelling (Aasland et al. 1996; Boyer et al. 2004). In addition, MSANTD was found to be associated with neuro-developmental diseases (Etchegaray et al. 2022) and cancer, including salivary gland acinic cell carcinoma (Barasch et al. 2017). Its inactivation in zebrafish caused development delays in tail and nervous system malformation (Etchegaray et al. 2022). Furthermore, the 19,564 G/C SNP marker was found to be in high association with zinc finger and SCAN domain-containing (ZSCAN) transcription factor. The complex regulation mechanism of the ZSCAN transcription factor allowed promotive or prohibitive effects on cancer progression, including cell apoptosis, cell migration and invasion, angiogenesis, cell differentiation, stem cell properties, cell proliferation, and chemotherapy sensitivity (Huang et al. 2019). Among the ZSCAN members, ZSCAN29 had shown the greatest range of pleiotropic effects among 80 SNPs with pleiotropic effects on psychiatric disorders, including its etiological role which influences DNA methylation and gene expression in the brain (Pineda-Cirera et al. 2022).

In this study, 6151 C/T and 27,857 A/T SNP had shown a single melting temperature (T_m) peak at 76.6 °C for allele C and 75.6 °C for allele T, respectively after subjecting three random samples from each group to Sanger sequencing. 112,388 G/A and 19,564 G/C SNP markers had shown homozygous (single T_m peak) and heterozygous (double T_m peaks). Both primers were showing amplicons with close T_m values as there were double T_m peaks produced for a heterozygous sample, with both of its T_m values similar to each allele (Baris et al. 2013). T-Plex real-time PCR assay had been applied in genotyping β -thalassaemia among Iraqi patients (Al-Abedy et al. 2019) and introns found on human genes TLR4 and S100 A9 (Dhas et al. 2015). Authors had summarized the assay as a swift and cost-effective assay to provide accurate allelic discrimination and zygosity testing despite some complexity in primer design (Dhas et al. 2015; Al-Abedy et al. 2019).

Successful marker-assisted selections (MAS) were reported among fish, particularly Atlantic salmon, with the discovery of three biomarkers to fight against infectious pancreatic necrosis and their application into MAS (Moen et al. 2009). Furthermore, later in the year 2017, a major quantitative trait locus associated with red sea bream iridoviral disease (RSIVD)-resistance was applied within the selection of red sea bream (Sawayama et al. 2017). Furthermore, MAS was reported to be able to reduce the disease outbreaks of infectious pancreatic necrosis (IPN) by 75% and it is further applied into breeding programs in Scotland and Norway to resist against IPN genetically (Houston et al. 2008; Moen et al. 2009). Despite its successful application in the aquaculture industry to reduce disease incidence, there were still limited production derived from the selectively bred stocks. However, the proportion was observed to be on the rise rapidly, specifically for high economic value species, including mollusks and crustaceans (Houston et al. 2020). The advancements in methodology and technology had allowed the evolution of selective breeding

from mass selection towards quantitative genetics and breeding methods, to candidate gene and quantitative trait loci (QTL) information and genomic selection. The introduction and advancement of genomics had accelerated genetic improvement programs and production systems. It allowed a more precise understanding of the relationships between animals, identifying markers associated with biological and economic significance, and enhancing the reliability of breeding value estimates, particularly for traits with low heritability (Houston et al. 2020). Nonetheless, in light of the dynamic interplay between hosts and pathogens, breeding for disease resistance may inadvertently drive pathogen evolution to overcome resistance mechanisms (Masri et al. 2003). Therefore, it is imperative for the continuous monitoring of the changes in pathogen characteristics, including molecular profiles, virulence, and antimicrobial resistance (Sciuto et al. 2022).

In a nutshell, with the successful identification and validation of 19,564 G/C SNP marker associated with disease resistance towards *Pseudomonas koreensis* in *Tor tambroides* via whole-genome pooled sequencing and T-plex ARMS PCR assay in this study, it is hoped that the marker can be used in molecular breeding for the disease-resistant trait of Malaysian mahseer.

Conclusion

In summary, disease-resistant single nucleotide polymorphisms (SNPs) associated with *Pseudomonas koreensis* were identified and validated through whole genome pooled sequencing, T-plex ARMS real-time PCR assay, and Sanger sequencing on resistant and susceptible fish pools. The successful validation of these SNP markers is expected to contribute to future studies on disease-resistant associations and can be incorporated into breeding programs to integrate genomic data into artificial selection decisions for Malaysian empurau breeding.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10499-025-02055-z>.

Author contributions M.M.L.L. : Writing-Original Draft, Data curation, Conceptualization, Funding Acquisition. H.H.C.: Conceptualization, Funding acquisition, Writing-Review and Editing. C.J.Y.K.: Data curation, Conceptualization. H.M.G.: Methodology, Conceptualization, Writing-Review and Editing. A.Z.: Data curation, Conceptualization.

Funding Open access funding provided by The Ministry of Higher Education Malaysia and Universiti Malaysia Sarawak. This research was fully supported by the Fundamental Research Grant Scheme with grant number FRGS/1/2022/STG01/UNIMAS/02/2 awarded to H. H. Chung by the Ministry of Higher Education Malaysia as well as the Universiti Malaysia Sarawak Postgraduate Student Research Grant with grant number UNI/F07/GRADUATES/85570/2023 awarded to M. M. L. Lau.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The

images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Aasland R, Stewart AF, Gibson T (1996) The SANT domain: a putative DNA binding domain in the SWI-SNF and ADA complexes, the transcriptional corepressor N-CoR and TFIIB. *Trends Biochem Sci* 3(21):87–88
- Al-Abedy NM, Salman ED, Faraj SA (2019) Frequency of human hemochromatosis HFE gene mutations and serum hepcidin level in iron overload β -thalassaemia Iraqi patients. *Public Health* 22(10):S275
- Al-Tobasei R, Ali A, Leeds TD, Liu S, Palti Y, Kenney B, Salem M (2017) Identification of SNPs associated with muscle yield and quality traits using allelic imbalance analyses of pooled RNA-Seq samples in rainbow trout. *BMC Genomics* 18:1–15
- Bajpai PK, Harel A, Shafir S, Barazani O (2022) Whole genome sequencing reveals footprints of adaptive genetic variation in populations of *Eruca sativa*. *Front Ecol Evol* 10:938981
- Baltrušis P, Doyle SR, Halvarsson P, Höglund J (2022) Genome-wide analysis of the response to ivermectin treatment by a Swedish field population of *Haemonchus contortus*. *Int J Parasitol Drugs Drug Resist* 18:12–19
- Bankers L, Fields P, McElroy KE, Boore JL, Logsdon JM Jr, Neiman M (2017) Genomic evidence for population-specific responses to co-evolving parasites in a New Zealand freshwater snail. *Mol Ecol* 26(14):3663–3675
- Barasch N, Gong X, Kwei K. A., Varma S, Biscocho J, ... & Pollack, J. R. (2017). Recurrent rearrangements of the Myb/SANT-like DNA-binding domain containing 3 gene (MSANTD3) in salivary gland acinic cell carcinoma. *PLOS One*, 12(2), e0171265.
- Baris I, Etlík O, Koksal V, Ocak Z, Baris ST (2013) SYBR green dye-based probe free SNP genotyping: introduction of T-Plex real-time PCR assay. *Anal Biochem* 441(2):225–231
- Boitard S, Schlötterer C, Nolte V, Pandey RV, Futschik A (2012) Detecting selective sweeps from pooled next-generation sequencing samples. *Mol Biol Evol* 29(9):2177–2186
- Bondad-Reantano, M. G., MacKinnon, B., Karunasagar, I., Fridman, S., Alday-Sanz, V., Brun, E., ... & Caputo, A. (2023). Review of alternatives to antibiotic use in aquaculture. *Reviews in Aquaculture*, 15(4), 1421–1451.
- Boyer LA, Latek RR, Peterson CL (2004) The SANT domain: a unique histone- tail binding module? *Nat Rev Mol Cell Biol* 5(2):158–163
- Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol* 38(12):5825–5829
- Carr IM, Robinson JI, Dimitriou R, Markham AF, Morgan AW, Bonthron DT (2009) Inferring relative proportions of DNA variants from sequencing electropherograms. *Bioinformatics* 25(24):3244–3250
- Chen J, Tian W (2016) Explaining the disease phenotype of intergenic SNP through predicted long range regulation. *Nucleic Acids Res* 44(18):8641–8654
- Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34(17):i884–i890
- Chiew, I., Salter, A. M., & Lim, Y. S. (2019). The significance of major viral and bacterial diseases in Malaysian aquaculture industry. *Pertanika Journal of Tropical Agricultural Science*, 42(3).
- Cho, S. C., Kim, J. W., Kim, B. N., Hwang, J. W., Shin, M. S., ... & Park, T. W. (2008). Association between the alpha-2C-adrenergic receptor gene and attention deficit hyperactivity disorder in a Korean sample. *Neuroscience Letters*, 446(2–3), 108–111.
- Collins A, Ke X (2012) Primer1: primer design web service for tetra-primer ARMS-PCR. *The Open Bioinformatics Journal* 6:55–58
- Dagnachew B, Aslam ML, Hillestad B, Meuwissen T, 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
- Danecek P, Bonfield J, Liddle J, Marshall J, Ohan V, Pollard M, Whitwham A, Keane T, McCarthy S, Davies R, Li H (2021) Twelve years of SAMtools and BCFtools. *Gigascience*, 10(2), giab008.

- Department of Fisheries (DOF). (2019) Annual fisheries statistic. Department of Fisheries Malaysia. Ministry of Agricultural and Agro-Based Industry, Putrajaya Retrieved from <https://www.dof.gov.my/index.php/pages/view/82>
- Department of Fisheries Malaysia (DOF) (2022). Perangkaan perikanan tahunan 2022. Retrieved from <https://ekkowunwkmj.exactdn.com/wp-content/uploads/2023/07/perangkaan-perikanan-2022-jilid-1-9.zip>
- Dhas DBB, Ashmi AH, Bhat BV, Parija SC, Banupriya N (2015) Modified low cost SNP genotyping technique using cycle threshold (Ct) & melting temperature (T_m) values in allele specific real-time PCR. *Indian J Med Res* 142(5):555–562
- Etchegaray, E., Baas, D., Naville, M., Haftek-Terreau, Z., & Volff, J. N. (2022). The neurodevelopmental gene MSANTD2 belongs to a gene family formed by recurrent molecular domestication of harbinger transposons at the base of vertebrates. *Molecular Biology and Evolution*, 39(8), msac173.
- FAO. (2016). The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all. Rome. Retrieved from <http://www.fao.org/3/i5555e/i5555e.pdf>
- Feng Y, Walsh CA (2004) The many faces of filamin: a versatile molecular scaffold for cell motility and signalling. *Nat Cell Biol* 6(11):1034–1038
- Fernandez BA, Scherer SW (2017) Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. *Dialogues Clin Neurosci* 9(4):353–371. <https://doi.org/10.31887/DCNS.2017.19.4/sscherer>
- Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., ... & Widmer, A. (2013). Population genomic footprints of selection and associations with climate in natural populations of *Arabidopsis halleri* from the Alps. *Molecular Ecology*, 22(22), 5594–5607.
- Fujita, M., Mitsuhashi, H., Isogai, S., Nakata, T., Kawakami, A., Nonaka, I., ... Kudo, A. (2012). Filamin C plays an essential role in the maintenance of the structural integrity of cardiac and skeletal muscles, revealed by the medaka mutant zacro. *Developmental Biology*, 361(1), 79–89.
- Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., ... Estoup, A. (2013). Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Molecular Ecology*, 22(14), 3766–3779.
- Geng X, Liu S, Yuan Z, Jiang Y, Zhi D, Liu Z (2017) A genome-wide association study reveals that genes with functions for bone development are associated with body conformation in catfish. *Mar Biotechnol* 19:570–578
- Guirao-Rico S, González J (2021) Benchmarking the performance of Pool-seq SNP callers using simulated and real sequencing data. *Mol Ecol Resour* 21(4):1216–1229
- Houston, R. D., Gheyas, A., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., ... & Bishop, S. C. (2008). Detection and confirmation of a major QTL affecting resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon (*Salmo salar*). In *Animal genomics for animal health* (Vol. 132, pp. 199–204). Karger Publishers.
- Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., ... & Robledo, D. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*, 21(7), 389–409.
- Huang M, Chen Y, Han D, Lei Z, Chu X (2019) Role of the zinc finger and SCAN domain-containing transcription factors in cancer. *Am J Cancer Res* 9(5):816
- Ingram B, Sungan S, Gooley G, Sim SY, Tinggi D, De Silva SS (2005) Induced spawning, larval development and rearing of two indigenous Malaysian mahseer, *Tor tambroides* and *Tor douronensis*. *Aquac Res* 36(10):983–995
- Irshath AA, Rajan AP, Vimal S, Prabhakaran VS, Ganesan R (2023) Bacterial pathogenesis in various fish diseases: recent advances and specific challenges in vaccine development. *Vaccines* 11(2):470
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240.
- Jumatli, A., & Ismail, M. S. (2021). Promotion of sustainable aquaculture in Malaysia. In *Proceedings of the International Workshop on the Promotion of Sustainable Aquaculture, Aquatic Animal Health, and Resource Enhancement in Southeast Asia*, 31–40.
- Kho CJY, Lau MML, Chung HH, Chew IYY, Gan HM (2023) Whole-genome sequencing of *Pseudomonas koreensis* isolated from diseased *Tor tambroides*. *Curr Microbiol* 80(8):255
- Khoo LW, Srinivasan SS, Henriquez FL, Bal AM (2021) A rare case of mixed infectious keratitis caused by *Pseudomonas koreensis* and *Aspergillus fumigatus*. *Case Reports in Ophthalmology* 11(3):600–605
- Klug A, Schwabe JW (1995) Zinc fingers. *FASEB J* 9(8):597–604
- Kofler R, Pandey RV, Schlötterer C (2011) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 27(24):3435–3436

- Konczal M, Koteja P, Stuglik MT, Radwan J, Babik W (2014) Accuracy of allele frequency estimation using pooled RNA-Seq. *Mol Ecol Resour* 14(2):381–392
- Kottelat, M., A. Pinder, and A. Harrison. (2018). *Tor tambroides*. The IUCN red list of threatened species 2018: e.T187939A91076554. Retrieved from <https://www.iucnredlist.org/species/187939/91076554>
- Lau, M. L. M., Kho, C. J. Y., Whye Kit Lim, L., Sia, S. C., Chung, H. H., Lihan, S., & Apun, K. (2022). Microbiome analysis of gut bacterial communities of healthy and diseased Malaysian mahseer (*Tor tambroides*) using 16S rRNA metagenomics approach. *Malaysian Journal of Microbiology*, 18(2).
- Lau, M. M. L., Kho, C. J. Y., Chung, H. H., & Zulkharnain, A. (2024). Isolation, identification and characterisation of *Pseudomonas koreensis* CM-01 isolated from diseased Malaysian mahseer (*Tor tambroides*). *Fish & Shellfish Immunology*, 109518.
- Lefkowitz RJ (2004) Historical review: a brief history and personal retrospective of seven transmembrane receptors. *Trends Pharmacol Sci* 25(8):413–422
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5):589–595
- Lionel, A. C., Crosbie, J., Barbosa, N., Goodale, T., Thiruvahindrapuram, B., Rickaby, J., ... & Scherer, S. W. (2011). Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Science Translational Medicine*, 3(95), 95ra75–95ra75.
- Lv J, Zhang D, Gao B, Liu P, Li J (2015) Transcriptome and MassARRAY analysis for identification of transcripts and SNPs for growth traits of the swimming crab *Portunus trituberculatus*. *Gene* 566(2):229–235
- Magnadottir B (2010) Immunological control of fish diseases. *Mar Biotechnol* 12:361–379
- Manyi-Loh C, Mamphweli S, Meyer E, Okoh A (2018) Antibiotic use in agriculture and its consequential resistance in environmental sources: potential public health implications. *Molecules* 23(4):795
- Masri, L., Branca, A., Sheppard, A. E., Papkou, A., Laehnemann, D., Guenther, P. S., ... & Go, S. J. (2003). *Pseudomonas koreensis* sp. nov., *Pseudomonas umsongensis* sp. nov. and *Pseudomonas jinjuensis* sp. nov., novel species from farm soils in Korea. *International Journal of Systematic and Evolutionary Microbiology*, 53(1), 21–27.
- Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., ... & Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease associated variation in regulatory DNA. *Science*, 337(6099), 1190–1195.
- Moen T, Baranski M, Sonesson AK, Kjølglum 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* 10:1–14
- Mohebiany, A. N., Harroch, S., & Bouyain, S. (2014). New insights into the roles of the contactin cell adhesion molecules in neural development. *Cell Adhesion Molecules: Implications in Neurological Diseases*, 165–194.
- Moss SM, Moss DR, Arce SM, Lightner DV, Lotz JM (2012) The role of selective breeding and biosecurity in the prevention of disease in penaeid shrimp aquaculture. *J Invertebr Pathol* 110(2):247–250
- Nakabayashi, K., Komaki, G., Tajima, A., Ando, T., Ishikawa, M., Nomoto, J., ... & Shirasawa, S. (2009). Identification of novel candidate loci for anorexia nervosa at 1q41 and 11q22 in Japanese by a genome-wide association analysis with microsatellite markers. *Journal of Human Genetics*, 54(9), 531–537.
- Nikpay M, Šeda O, Tremblay J, Petrovich M, Gaudet D, Kotchen TA, Cowley AW, Hamet P (2012) Genetic mapping of habitual substance use, obesity-related traits, responses to mental and physical stress, and heart rate and blood pressure measurements reveals shared genes that are overrepresented in the neural synapse. *Hypertens Res* 35(6):585–591
- Njeim R, Eid A (2018) Alpha-2c adrenergic receptor promotes the malignant phenotype of colon cancer cells. *FASEB J* 32:695–705
- Novotny L, Dvorska L, Lorencova A, Beran V, Pavlik I (2004) Fish: a potential source of bacterial pathogens for human beings. *Vet Med* 49(9):343–358
- Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., ... & Nishimura, Y. (1994). Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, 79(4), 639–648.
- Oguro-Ando A, Zuko A, Kleijer KT, Burbach JPH (2017) A current view on contactin-4, -5, and -6: implications in neurodevelopmental disorders. *Mol Cell Neurosci* 81:72–83
- Pękała-Safińska A (2018) Contemporary threats of bacterial infections in freshwater fish. *Journal of Veterinary Research* 62(3):261
- Pineda-Cirera L, Cabana-Dominguez J, Lee PH, Fernandez-Castillo N, Cormand B (2022) Identification of genetic variants influencing methylation in brain with pleiotropic effects on psychiatric disorders. *Prog Neuropsychopharmacol Biol Psychiatry* 113:110454

- Rellstab C, Zoller S, Tedder A, Gugerli F, Fischer MC (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
- Ren, S., Lyu, G., Irwin, D. M., Liu, X., Feng, C., Luo, R., ... & Wang, Z. (2021). Pooled sequencing analysis of geese (*Anser cygnoides*) reveals genomic variations associated with feather color. *Frontiers in Genetics*, 12, 650013.
- Ribeiro da Cunha B, Fonseca LP, Calado CR (2019) Antibiotic discovery: where have we come from, where do we go? *Antibiotics* 8(2):45
- Rivero G, Martín-Guerrero I, de Prado E, Gabilondo AM, Callado LF, García Sevilla JA, Garcia-Orad A, Meana JJ (2016) Alpha2C-adrenoceptor Del322-325 polymorphism and risk of psychiatric disorders: significant association with opiate abuse and dependence. *The World Journal of Biological Psychiatry* 17(4):308–315
- Rohani, M. F., Islam, S. M., Hossain, M. K., Ferdous, Z., Siddik, M. A., Nuruzzaman, M., ... & Shahjahan, M. (2022). Probiotics, prebiotics and synbiotics improved the functionality of aquafeed: upgrading growth, reproduction, immunity and disease resistance in fish. *Fish & Shellfish Immunology*, 120, 569–589.
- Sawayama E, Tanizawa S, Kitamura SI, Nakayama K, Ohta K, Ozaki A, Takagi M (2017) Identification of quantitative trait loci for resistance to RSIVD in red sea bream (*Pagrus major*). *Mar Biotechnol* 19:601–613
- Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., ... & Sherry, S. T. (2021). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 49(D1), D10.
- Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals—mining genome-wide polymorphism data without big funding. *Nat Rev Genet* 15(11):749–763
- Sciuto, S., Colli, L., Fabris, A., Pastorino, P., Stoppani, N., Esposito, G., ... & Colussi, S. (2022). What can genetics do for the control of infectious diseases in aquaculture?. *Animals*, 12(17), 2176.
- Shahi N, Mallik SK (2014) Recovery of *Pseudomonas koreensis* from eye lesions in golden mahseer, *Tor putitora* (Hamilton, 1822) in Uttarakhand, India the *Journal of Fish Disease* 37(5):497–500
- Shaul O (2017) How introns enhance gene expression. *Int J Biochem Cell Biol* 91:145–155
- Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 33(suppl_2), W465–W467.
- Stossel TP, Condeelis J, Cooley L, Hartwig JH, Noegel A, Schleicher M, Shapiro SS (2001) Filamins as integrators of cell mechanics and signalling. *Nat Rev Mol Cell Biol* 2(2):138–145
- Tahirov TH, Sasaki M, Inoue-Bungo T, Fujikawa A, Sato K, Kumasaka T, Yamamoto M, Ogata K (2001) Crystals of ternary protein–DNA complexes composed of DNA binding domains of c-Myb or v-Myb, C/EBP α or C/EBP β and tom-1A promoter fragment. *Acta Crystallogr D Biol Crystallogr* 57(11):1655–1658
- Tan, X., Cao, F., Tang, F., Lu, C., Yu, Q., Feng, S., ... & Sun, L. (2021). Suppression of DLBCL progression by the E3 ligase Trim35 is mediated by CLOCK degradation and NK cell infiltration. *Journal of Immunology Research*, 2021(1), 9995869.
- Uddin, T. M., Chakraborty, A. J., Khusro, A., Zidan, B. R. M., Mitra, S., Emran, T. B., ... & Koirala, N. (2021). Antibiotic resistance in microbes: history, mechanisms, therapeutic strategies and future prospects. *Journal of Infection and Public Health*, 14(12), 1750–1766.
- Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genet Sel Evol* 34(3):275–305
- Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., Marchler, G.H., ... & Marchler-Bauer, A. (2023). The conserved domain database in 2023. *Nucleic Acids Research*, 51(D1), D384–D388.
- Watkins RR, Bonomo RA (2016) Overview: global and local impact of antibiotic resistance. *Infect Dis Clin* 30(2):313–322
- Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., ... & Yaschenko, E. (2007). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 36(suppl_1), D13–D21.
- World Bank. (2013). Fish to 2030: prospects for fisheries and aquaculture (English). Retrieved from <http://documents.worldbank.org/curated/en/45863146815237668/pdf/831770WPOP11260ES003000Fish0to2030.pdf>
- Ye S, Dhillon S, Ke X, Collins AR, Day IN (2001) An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res* 29(17):e88–e88

- Yue X, Wang H, Huang X, Wang C, Chai X, Wang C, Liu B (2012) Singlenucleotide polymorphisms in i-type lysozyme gene and their correlation with *Vibrio* resistance and growth of clam *Meretrix meretrix* based on the selected resistance stocks. *Fish Shellfish Immunol* 33(3):559–568
- Zhang Q, Yu Y, Wang Q, Liu F, Luo Z, Zhang C, Zhang X, Huang H, Xiang, J., & Li, F. (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* 10:700
- Zhao X, Wan J, Fu J, Shao Y, Lv Z, Li C (2021) Identification of SNPs associated with disease resistance in juveniles of *Sinonovacula constricta* using RNA-seq and high-resolution melting analysis. *Aquaculture* 544:737109
- Zilberberg, M. D., & Shorr, A. F. (2012). Economics at the end of life: hospital and ICU perspectives. In *Seminars in Respiratory and Critical Care Medicine*, 33(4), 362369. Thieme Medical Publishers.

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

Authors and Affiliations

Melinda Mei Lin Lau¹ · Hung Hui Chung¹ · Cindy Jia Yung Kho² · Han Ming Gan³ · Azham Zulkharnain⁴

✉ Hung Hui Chung
hhchung@unimas.my

Melinda Mei Lin Lau
mlaumeilin@gmail.com

Cindy Jia Yung Kho
cindy.khojy@ntu.edu.sg

Han Ming Gan
gan.mypb@gmail.com

Azham Zulkharnain
azham@shibaura-it.ac.jp

¹ Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

² School of Biological Sciences, Nanyang Technological University, Singapore, Singapore

³ Patriot Biotech Sdn. Bhd., 47500 Bandar Sunway, Selangor, Malaysia

⁴ Department of Bioscience and Engineering, College of System Engineering and Science, Shibaura Institute of Technology, 307 Fukasaku, Minuma-Ku, Saitama 337-8570, Japan