



Faculty of Resource Science and Technology

**Genomic Profiling of Empurau (*Tor tambroides*) for the Discovery of
Single Nucleotide Polymorphisms (SNPs) Associated with Resistance to
*Pseudomonas koreensis***

Melinda Lau Mei Lin

**Doctor of Philosophy
2024**

Genomic Profiling of Empurau (*Tor tambroides*) for the Discovery of Single
Nucleotide Polymorphisms (SNPs) Associated with Resistance to
Pseudomonas koreensis

Melinda Lau Mei Lin

A thesis submitted

In fulfillment of the requirements for the degree of Doctor of Philosophy

(Molecular Genetics)

Faculty of Resource Science and Technology
UNIVERSITI MALAYSIA SARAWAK

2024

DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Malaysia Sarawak. Except where due acknowledgements have been made, the work is that of the author alone. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.



.....

Signature

Name: Melinda Lau Mei Lin

Matric No.: 22010005

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak

Date : 11/10/24

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my supervisor, Associate Professor Dr. Chung Hung Hui, for his unwavering support, guidance, and encouragement throughout my PhD journey. His expertise, mentorship, and constructive feedback have been invaluable in shaping the direction of my research. I also extend my heartfelt appreciation to my co-supervisor, Dr. Elexson Nillian, for his valuable insights and contributions to my research.

To my labmates, Ivy Chew and Cindy Kho, I am indebted for their camaraderie, collaboration, and friendship throughout this research endeavor. Their encouragement and willingness to share knowledge and resources have enriched my academic experience. Additionally, I am grateful to my close friends, Adrian Ling, Chee Hou, Sing Ying, Irene Hii, Adrianne Vong and Siew Chuiang for their unwavering support and understanding, which have been a source of strength during challenging times.

Furthermore, I would like to extend my deepest gratitude to my family — my father Lau Sieh Liung, my mother Ting Hung Yung, and brother Jasper Lau Yueh Hong — for their unwavering love, support, and sacrifice. My family's unconditional love, encouragement, and belief in my abilities have been a constant source of strength and motivation. I am especially grateful to my late mother, who is no longer with us, for her boundless love, wisdom, and guidance. Though she is no longer physically present, her memory and spirit continue to inspire and drive me forward in pursuit of my goals.

ABSTRACT

Empurau (*Tor tambroides*) is highly sought for its high nutritional value and unique flesh taste, allowing it to earn a top place among the most expensive freshwater fishes among aquaculture industry in Malaysia. Uncovering and characterising its genomic data is crucial for unmasking ambiguous taxonomy within *Tor* genus, while resolving its species identification for conservation purposes. A combination of Nanopore long reads and Illumina short reads had reported on an estimated size of 1.23 Gb with 87.6 % completeness of *T. tambroides* genome. The genomic data had been mapped to various pathways and showed significant similarity to published immune- and growth-related genes. The causative agent of the diseased empurau at the local aquaculture farm was identified as *Pseudomonas koreensis* with its virulence, antibiotic sensitivity and growth characteristics being characterised. It displays notable virulence characteristics that led to the acute infection of the fish. The isolated strain, CM-01, had been characterised on its virulence genes, antibiotics susceptibility testing and growth. Experimental infection had revealed the lethal dosage of CM-01 strain as 5.01×10^7 CFU/mL along with significant histological lesions observed. Its pathogenicity was found to be correlated with twelve virulence genes identified which encodes for iron uptake, protease, acylhomoserine lactone synthase gacS/gacA component regulation system, type IV secretion system, hydrogen cyanide production, exolysin, alginate biosynthesis, flagella and pili. *P. koreensis* was tested susceptible to beta-lactam combination agents which further suggested as its drug of choice. Its growing features had shown its optimal growth rate at pH 7 and 30 °C. These findings could establish a scientific foundation for the recognition of the pathogen and fish management. Given antibiotics administration as the common solution to disease outbreaks, contributing to the worsening of antimicrobial resistance (AMR), the development of disease-resistant

broodstock through the identification of SNP markers appeared as one of a promising strategy. The SNP markers can be identified via whole genome pooled sequencing on samples treated with LD₅₀, forming resistant and susceptible group. Multiple tests had been 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 had been performed to validate the significance of the candidate marker. With the successful validation of the 19564 G/C SNP marker, it is hoped to contribute to further studies considering disease-resistant association SNPs and can be used into breeding programs to integrate genomic information into artificial selection decisions in empurau breeding.

Keywords: Empurau genome, *Pseudomonas koreensis*, disease-resistant SNP markers, antibiotics resistance in fish, aquaculture disease management

Profil Genomik empurau (Tor tambroides) untuk Penemuan Polimorfisme Nukleotida Tunggal (SNP) yang dikaitkan dengan Rintangan terhadap Pseudomonas koreensis

ABSTRAK

Empurau (Tor tambroides) sangat dicari kerana nilai pemakanan yang tinggi dan rasa daging yang unik, membolehkan ia mendapat tempat teratas dalam kalangan ikan air tawar paling mahal dalam kalangan industri akuakultur di Malaysia. Mendedahkan dan mencirikan data genomnya adalah penting untuk membuka taksonomi samar-samar dalam genus Tor, sambil menyelesaikan pengenalan spesiesnya untuk tujuan pemuliharaan. Gabungan bacaan panjang Nanopore dan bacaan pendek Illumina telah melaporkan anggaran saiz 1.23 Gb dengan 87.6 % kesempurnaan genom T. tambroides. Data genomik telah dipetakan ke pelbagai laluan dan menunjukkan persamaan yang ketara dengan gen berkaitan imun dan pertumbuhan yang diterbitkan. Ejen penyebab empurau berpenyakit di ladang akuakultur tempatan dikenal pasti sebagai Pseudomonas koreensis dengan ciri virulensi, sensitiviti antibiotik dan ciri pertumbuhannya. Ia memaparkan ciri virulensi yang ketara yang membawa kepada jangkitan akut ikan. Strain terpencil, CM-01, telah dicirikan pada gen virulensinya, ujian kerentanan antibiotik dan pertumbuhannya. Jangkitan eksperimen telah mendedahkan dos maut strain CM-01 sebagai 5.01×10^7 CFU/mL bersama-sama dengan lesi histologi yang ketara diperhatikan. Kepatogenannya didapati berkorelasi dengan dua belas gen virulensi yang dikenal pasti yang mengodkan pengambilan besi, protease, acylhomoserine lactone synthase gacS/gacA komponen sistem, sistem rembesan jenis IV, pengeluaran hidrogen sianida, exolisin, biosintesis alginat, flagela dan pili. P. koreensis telah diuji terdedah kepada agen gabungan beta-laktam yang selanjutnya dicadangkan sebagai ubat pilihannya. Ciri-ciri pertumbuhannya telah menunjukkan kadar pertumbuhan optimumnya pada pH 7 dan 30 °C. Penemuan ini boleh

mewujudkan asas saintifik untuk pengiktirafan patogen dan pengurusan ikan. Memandangkan pentadbiran antibiotik sebagai penyelesaian biasa kepada wabak penyakit, menyumbang kepada kemerosotan rintangan antimikrobial (AMR), pembangunan stok induk tahan penyakit melalui pengenalpastian penanda SNP muncul sebagai salah satu strategi yang menjanjikan. Penanda SNP boleh dikenal pasti melalui penjujukan terkumpul genom keseluruhan pada sampel yang dirawat dengan LD₅₀, membentuk kumpulan yang tahan dan mudah terdedah. Pelbagai ujian telah dijalankan untuk menyaring dan memilih penanda SNP calon untuk pengesahan selanjutnya menggunakan ujian PCR masa nyata T-plex ARMS. Keputusan genotaip pada penanda calon yang dipilih telah disahkan menggunakan penjujukan Sanger. Analisis statistik telah dilakukan untuk mengesahkan kepentingan penanda calon. Dengan kejayaan pengesahan penanda SNP G/C 19564, diharapkan dapat menyumbang kepada kajian lanjut yang mempertimbangkan SNP persatuan yang tahan penyakit dan boleh digunakan ke dalam program pembiakan untuk mengintegrasikan maklumat genomik ke dalam keputusan pemilihan buatan dalam pembiakan empurau.

Kata kunci: Genom empurau, Pseudomonas koreensis, penanda SNP tahan penyakit, rintangan antibiotik dalam ikan, pengurusan penyakit akuakultur

TABLE OF CONTENTS

	Page
DECLARATION	i
ACKNOWLEDGEMENT	ii
ABSTRACT	iii
ABSTRAK	v
TABLE OF CONTENTS	vii
LIST OF TABLES	xiii
LIST OF FIGURES	xv
LIST OF ABBREVIATIONS	xviii
CHAPTER 1: INTRODUCTION	1
1.1 Current Status of Aquaculture in Malaysia	1
1.2 Fish Genomes and its Significance	3
1.3 <i>Tor tambroides</i> Overview	5
1.4 Taxonomical Status of <i>Tor</i>	7
1.5 Aquaculture Diseases	9
1.6 <i>Pseudomonas</i> Pathogenicity and Prevalence in Environmental and Clinical Settings	11
1.7 Ontogeny of the Immune System	12
1.8 Fish Immunity	14

1.8.1	Innate Component	17
1.8.2	Adaptive Component	19
1.9	Factors Affecting Fish Immune System	21
1.9.1	Inherent Factors	21
1.9.2	External Factors	23
1.10	Pathogen-Host Association	25
1.11	Antimicrobial Resistance (AMR) and Factors Attributing to it	27
1.11.1	Microbial (Natural) Factors	28
1.11.2	Antibiotics Misuse and Overuse	29
CHAPTER 2: EMPURAU (<i>T. TAMBROIDES</i>) GENOME SEQUENCING AND ITS PHYLOGENETICS		33
2.1	Introduction	33
2.2	Materials and Methods	36
2.2.1	Ethics Statement	36
2.2.2	Sampling and DNA Extraction	36
2.2.3	Species Verification	36
2.2.4	Library Construction and Whole Genome Sequencing	37
2.2.5	Sequence Data Processing and Assembly	37
2.2.6	Detection of Repetitive Sequences and Prediction of Protein-Coding Genes	38
2.2.7	SSR Analysis	38

2.2.8	Functional Annotation of Protein-Coding Genes	39
2.2.9	Ortholog Interference and Phylogenetic Construction	39
2.3	Results	44
2.3.1	Characterisation of <i>Tor tambroides</i> Genome	44
2.3.2	Repeat-Content Analysis	46
2.3.3	SSR Analysis	49
2.3.4	Functional Annotation of <i>T. tambroides</i>	51
2.3.5	Immunity-Related Annotation	59
2.3.6	Growth-related Annotation	63
2.3.7	Orthologs and Phylogenetic Interferences	70
2.4	Discussions	75
2.4.1	SSR Analysis	76
2.4.2	Accumulation of Repetitive Elements in Shaping <i>Tor tambroides</i> Genome	78
2.4.3	Immunity-Related Annotation	81
2.4.4	Growth-Related Annotation	84
 CHAPTER 3: ISOLATION, IDENTIFICATION AND CHARACTERISATION		
 OF <i>P. KOREENSIS</i> ISOLATED FROM DISEASED EMPURAU		
 (<i>T. TAMBROIDES</i>)		
		86
3.1	Introduction	86
3.2	Materials and Methods	89

3.2.1	Bacterial Isolation	89
3.2.2	Physiological and Biochemical Characterisation	89
3.2.3	16S rRNA Gene Sequence Analysis	90
3.2.4	Growing Characteristics	90
3.2.5	Drug Sensitivity Tests	91
3.2.6	Screening of Virulence Genes	91
3.2.7	Experimental Infections	94
3.2.8	Histopathological Observations	94
3.3	Results	96
3.3.1	Morphological and Biochemical Characteristics	96
3.3.2	Sequence Analysis of the 16S rRNA Gene	98
3.3.3	Experimental Infections	99
3.3.4	Histopathological Observations	102
3.3.5	Drug Sensitivity Test	105
3.3.6	Virulence Genes	106
3.3.7	Growing Characteristics	107
3.4	Discussion	109
3.4.1	Characterisation of <i>Pseudomonas koreensis</i> CM-01	110
3.4.2	Experimental Challenge	111
3.4.3	Virulence and Toxicity	114

3.4.4	Antibiotics Susceptibility	117
-------	----------------------------	-----

CHAPTER 4: IDENTIFICATION OF SINGLE NUCLETOIDE

POLYMORPHISMS (SNPs) ASSOCIATED WITH DISEASE

RESISTANCE TOWARDS *P. KOREENSIS* IN EMPURAU

(*T. TAMBROIDES*) USING T-PLEX REAL-TIME PCR ASSAY 120

4.1	Introduction	120
4.2	Materials and Methods	124
4.2.1	Bacterial Challenge and Sample Preparation	124
4.2.2	Sequencing and Read Cleaning	124
4.2.3	Reads mapping, SNP Calling and Identification	125
4.2.4	T-plex Real-Time PCR Assay	126
4.2.5	Primer Design	128
4.2.6	SNP Validation	131
4.3	Results	132
4.3.1	General Overview and SNP Identification	132
4.3.2	Functional Annotation	135
4.3.3	SNP Validation via T-Plex Real-Time PCR Assay	140
4.4	Discussions	146
4.4.1	Marker Assisted Selection as Antimicrobial Resistance Solution	146
4.4.2	T-Plex ARMS PCR Assay & Pooled Sequencing	149

4.4.3	Identification of Disease-Resistant SNPs against <i>Pseudomonas koreensis</i>	151
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS		157
5.1	Conclusions	157
5.2	Recommendations	158
REFERENCES		159
APPENDICES		219

LIST OF TABLES

	Page
Table 1.1	Maturation of immunological competence in carps 14
Table 2.1	Available genomes of fish species and their genome statistics used for the Maximum-likelihood tree construction 41
Table 2.2	Genomic sequencing and assembly statistic of <i>Tor tambroides</i> 44
Table 2.3	Summary of repeats within the genome of <i>T. tambroides</i> 48
Table 2.4	The number, total length and average length of five different types of SSRs found within the genome of <i>T. tambroides</i> 50
Table 2.5	Functional annotation of predicted protein sequences to the various database 52
Table 2.6	Immune-related genes 61
Table 2.7	Growth-related genes 65
Table 2.8	BUSCO completeness of 52 ray-finned fishes for the inference of phylogenetic relationship with <i>T. tambroides</i> using BUSCO supermatrix approach 71
Table 3.1	PCR reaction components for virulence genes amplification from <i>P. koreensis</i> 92
Table 3.2	PCR targeted virulence genes, targeted amplicon size and annealing temperatures of CM-01 strain 93
Table 3.3	Mortality of fish infected with CM-01 isolate across 7 days experimental observation with the p-value shown from Tukey test in comparison with control group 100
Table 3.4	Antibiotics sensitivity of CM-01 to different drugs 105

Table 4.1	Primer sequences for potential SNP validation designed using a web-based software	129
Table 4.2	Summary statistics for both resistant group and susceptible group from pooled sequencing	132
Table 4.3	SNP selection including CMH test, pairwise F_{ST} and FE test were performed for further validation using T-plex real-time PCR assay	134
Table 4.4	Functional annotations of the five selected SNP markers	137
Table 4.5	Distribution of the five markers in the susceptible and resistant groups within <i>T. tambroides</i>	140

LIST OF FIGURES

	Page
Figure 1.1 Productions by subsectors expected from year 2021 to 2030 by Department of Fisheries Malaysia with reference to Dasar Agromakanan Negara (DAN) 2.0	2
Figure 1.2 Sketching of Empurau (<i>Tor tambroides</i>).	7
Figure 1.3 Reconstruction of maximum likelihood phylogenetic tree utilizing 13 protein-coding genes and two ribosomal genes across 16 <i>Tor</i> species with 1000 bootstrap	9
Figure 1.4 Functional scope and ecological niches of genus <i>Pseudomonas</i> , highlighting on the extensive distribution of <i>P. fluorescens</i> species complex with addition of <i>Pseudomonas koreensis</i>	12
Figure 1.5 The origin and evolutionary connections between innate and adaptive immunity	16
Figure 2.1 Estimation of genome size, repeat content and heterozygosity by GenomeScope	45
Figure 2.2 Length distribution comparison on total gene, exon, CDS and protein of annotated genes of <i>Tor tambroides</i> to <i>Danio rerio</i> , <i>Carassius auratus</i> , <i>Pimephales promelas</i> , <i>Sinocyclocheilus anshuiensis</i> , <i>Sinocyclocheilus grahami</i> and <i>Sinocyclocheilus rhinicerous</i>	46
Figure 2.3 SSRs percentage graph with selected top six frequency SSRs from each group	50
Figure 2.4 Venn diagram showing differences and similarity of predicted protein sequences of <i>T. tambroides</i> annotated to GO, KEGG, and COG databases	51
Figure 2.5 GO classifications of <i>T. tambroides</i> genome into biological processes, molecular function and cellular component categories	53
Figure 2.6 KEGG classifications of <i>T. tambroides</i> genome into organismal system, cellular processes, environmental information processing, genetic information processing and metabolism categories	54
Figure 2.7 Top 10 KEGG classification of <i>T. tambroides</i> genome into five main categories: organismal system, cellular processes, environmental information processing, genetic information processing and metabolism categories	57

Figure 2.8	COG classifications of <i>T. tambroides</i> genome into four main categories: poorly characterized, metabolism, cellular processes & signaling and information storage & processing	58
Figure 2.9	Annotation of <i>T. tambroides</i> genome to known immune-pathway within GO databases	60
Figure 2.10	Annotation of <i>T. tambroides</i> genome to known top 20 growth-related pathway in KEGG database	64
Figure 2.11	Maximum-likelihood tree plotted on 52 Actinopterygii species with one Sarcopterygii species as outgroup inferred from a supermatrix of 3,640 BUSCOs	74
Figure 3.1	Growth of CM-01 strain on different agar after culturing at 30 °C for 24 hours	97
Figure 3.2	The Gram Staining properties of CM-01 strain with pink appearance observed	97
Figure 3.3	Agarose gel electrophoresis of PCR products of 16S rRNA gene of CM-01 with the size estimated through 100 bp DNA ladder (Transgene, French)	98
Figure 3.4	Neighbour-joining phylogenetic tree based on 16S rRNA gene of <i>Pseudomonas</i> strains using 1,000 bootstraps	99
Figure 3.5	Log dose versus probit for the calculation of median lethal dosage of <i>P. koreensis</i> CM-01 in juvenile <i>T. tambroides</i>	101
Figure 3.6	Pathological features of <i>Tor tambroides</i> infected with CM-01 strain	101
Figure 3.7	Photomicrograph of spleen, intestine, gills and kidneys of control and infected empurau	104
Figure 3.8	Agarose gel electrophoresis visualizing the amplification products of virulence genes	106
Figure 3.9	Growing characteristics of <i>Pseudomonas koreensis</i> , CM-01 strain	108
Figure 4.1	Introduction of T-Plex real-time PCR assay	127
Figure 4.2	Classification of SNP identification into transition and transversion types identified within the genome of <i>T. tambroides</i>	132
Figure 4.3	The results of Sanger sequencing and T-plex real-time PCR analysis on 6151 C/T, 112388 G/A, 27857 A/T and 19564 G/C SNP	142

Figure 4.4 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 19564 G/C SNP

144

LIST OF ABBREVIATIONS

AMR	Antimicrobial Resistance
APC	Antigen Presenting Cell
COG	Clusters of Orthologous Genes
COX I / COI	Cytochrome Oxidase Subunit I
DNA	Dideoxynucleotide Acid
DOF	Department of Fisheries
FAO	Food and Agriculture Organization
GO	Gene Ontology
Ig	Immunoglobulin
IUCN	International Union for Conservation of Nature
KEGG	Kyoto Encyclopedia of Genes and Genomes
MAS	Marker-Assisted Selection
MHC	Major Histocompatibility Complex
PAMP	Pattern Associated Molecular Pattern
PRR	Pattern Recognition Receptor
QTL	Quantitative Trait Loci
SNP	Single Nucleotide Polymorphisms
SSR	Simple Sequence Repeats

CHAPTER 1

INTRODUCTION

1.1 Current Status of Aquaculture in Malaysia

Aquaculture plays a crucial role as a source of animal protein in diets worldwide (FAO, 2016; DOF, 2019), being the most affordable option for animal-based proteins (Jumatli & Ismail, 2021). By the year 2030, the world fish production is predicted to increase to 187 million metric tons (MT) with the prediction of stagnant capture fisheries and progressive increment of aquaculture industry until an equilibrium is reached to cope with the increasing world's population (World Bank, 2013). Thus, aquaculture production alone needs to increase by an additional of 46.4 million metric tons for sustainable food production and security (World Bank, 2013). The aquaculture industry can contribute to the national economy, particularly its Gross Domestic Product (GDP) (SEAFDEC, 2017).

Malaysia Fisheries sector including capture fisheries, inland fisheries and aquaculture produced up to 5.7 million tones with an estimated value of RM4 billion with an increment of 37.5 % and 17.2 % in quantity and value, respectively as compared to year 2021 (DOF, 2022). This reflects strong consumer demand for fish products, with per capita consumption reaching 59 kg in 2016, one of the highest globally (FAO, 2020a). Malaysia relies significantly on its seafood production sector, with aquaculture accounting for up to 20% of the total seafood production (FAO, 2020b). The Government of Malaysia has committed to enormous allocation of financial and physical services for the growth of aquaculture industry since 2003, particularly Aquaculture Industrial Zone (AIZ) projects, National Agro-Food Policy (NAFP 2011-2020) and Malaysia National Key Economics Area (NKEA) (Jumatli & Ismail, 2021). Over time, the industry has expanded steadily, reaching

an output of 1.69 million tons in 2017, following consistent production growth since 1980 (Waiho et al., 2020).

DOF had expected doubled aquaculture production within 10-year timeframe till year 2030 as it serves as a potential industry to be developed (Figure 1.1) to cope with the expanding human population. Aquaculture production had reported 573,682 metric tonnes (MT) worth RM 4 billion in year 2022, contributing 30.4% of the nation fisheries statistics (DOF, 2022). As compared to previous year, it showed an increment of 37.5% in quantity and 17.2% in value, further depicting its great underlying potential. In addition, Sarawak had ranked in top 5 across the country in the fisheries production in both quantity and value (DOF, 2022).

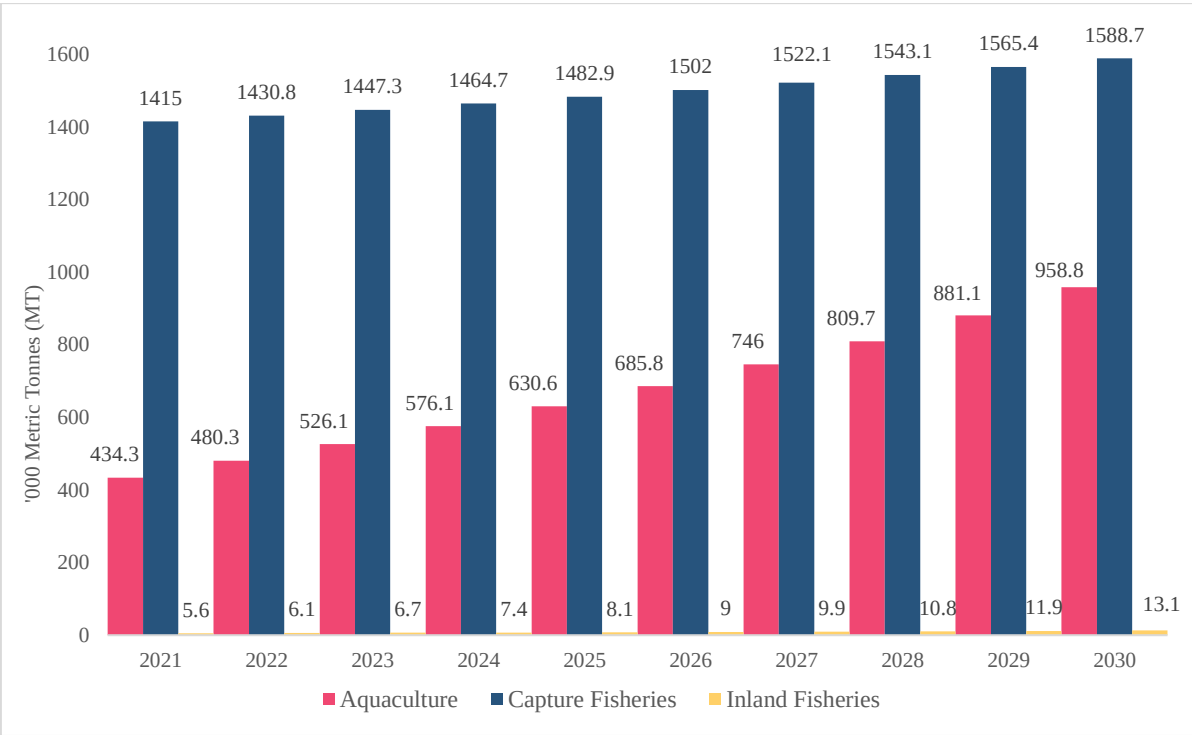


Figure 1.1: Productions by subsectors expected from year 2021 to 2030 by the Department of Fisheries Malaysia with reference to Dasar Agromakanan Negara (DAN) 2.0. The increase in the expected production of Aquaculture as well as Capture Fisheries with its focus on deep-sea fisheries as both industries lies a great development potential within 10-year period. Adapted from DOF, 2022

1.2 Fish Genomes and its Significance

Fishes are recognized as the largest and oldest category of vertebrates due to their extensive evolution history, possibly attributed to repeated rounds of whole genome duplication and re-diploidization, leading to increase genome complexity. Unlike mammalian genomes, teleost genomes contain multiple families of active transposable elements, which may have influenced speciation by affecting hybrid sterility and viability. Moreover, the diverse array of sex determination systems and adaptable sex chromosomes observed in teleost may contribute to pre- and post-mating reproductive isolation, resulting in more complex karyotypes and genome sizes as compared to other vertebrates (Volf, 2005). As the world population is increasing, the demand towards animal protein, especially fish production is on the rise, causing challenges in achieving sustainable food security and production. Sustainability efforts face obstacles such as overfishing, degradation of genetic resources, climate change and diseases (FAO, 2018). The establishment of sequencing technologies and genomic advances could play a crucial role in addressing some of these challenges and further benefiting sustainable fisheries and aquaculture (Lu & Luo, 2020).

The world took 24 years to develop technologies to sequence DNA since the first discovery of double helix DNA structure in year 1953. With the official launch of human whole genome sequencing in the mid-1980s, major changes were observed within scientific research with the switch from hypothesis-driven research to data-driven research, involving simple facts deduction from complex data (FAO, 2021). Aquaculture genomics began in the 1990s, with the first Aquaculture Genomics Workshop in 1997 in the United States, focusing on genomic research on six species (salmonids, tilapia, catfish, oysters, shrimps, striped bass). The intervention of genomics brought unity to all research fields of basic biology, including physiology, genetics and genomics, reproduction, nutrition and growth.

Advancements in Polymerase Chain Reaction and nanotechnologies propelled sequencing technologies, making DNA sequencing faster, cheaper and more accessible for scientific research and personalized medicine (FAO, 2021).

In the early 2000's, the first fish, Japanese pufferfish *Fugu rubripes* had its whole genome being sequenced, which is a commercial food fish in Japan and China (Aparicio et al., 2002). With the emergence of high-throughput sequencing technologies around 2005, more than two hundred fish genomes were successfully sequenced and made available in public repositories (Heather & Chain, 2006). Fisheries genomics became crucial not only in fields like comparative genomics, evolution, and systematics but also in aquaculture and fisheries, including breeding, disease management, maturation and sex determination (Yue & Wang, 2017). Nonetheless, as compared to agriculture, aquaculture and fisheries lagged in applying genomic information for selective breeding and resource management (Bernatchez et al., 2017).

The rapid advancement of DNA sequencing technologies facilitated a collaborative genomic study, aiming to map the vertebrate tree of life. Fish, consisting of more than half of all the vertebrates, initiated the Fish10K Genome Project, a sub-project of the Earth BioGenome Project, targeting to sequence and analyze the genomes of 10,000 fishes (Fan et al., 2020). Accelerated large-scale fish genome sequencing is expected to enhance research on fish biodiversity, speciation, conservation, adaptation and sustainable fish use (Fan et al., 2020).

Various applications were observed within aquaculture genomics, including preservation of genetic diversity in fish populations, fisheries management facilitation, enhancement of disease resistance, genome-based selection for genetic improvement as well as sex determination and regulation (Lu & Luo, 2020). Discovery of a draft genome can act

as a genomic resource for the study and facilitation of economically important traits and genetic breeding (Xu et al., 2014). Furthermore, genome analysis could serve as a basis for identification and better understanding of immune system and growth performances to improve on the disease resistance and growth traits (Lu & Luo, 2020). In addition, sex-related gene identification can facilitate marker-assisted selection as sex plays playing significant role in growth, reproduction and product quality (Budd et al., 2015; Figueras et al., 2016). The whole genome sequencing is able to resolve the population structure of a species to further encourage genetic diversity and environmental adaptation analysis (Vij et al., 2016; Xu et al., 2017).

1.3 *Tor tambroides* Overview

In recent years, like other *Tor* spp., the natural habitat of *T. tambroides* is facing environmental degradation which cause a significant drop in its population size (Ingram et al., 2005). River modifications due to human society demands including agricultural activities, logging and deforestation would disrupt and eventually lessen water flow in its habitat. Eventually, these activities apart from diminishing the population of *T. tambroides* and other *Tor* species, further bring to repercussions to aquatic environment too. Notwithstanding environmental concerns, *T. tambroides* also face dangers from excessive fishing activities due to the utilization of nets, hooks and dynamites and there is still a lack of information on such losses (Kottelat et al., 2018).

Due to its unique flesh taste enriched with sweetness, it is one of the prized species targeted for human consumption as well which is directly sold to restaurants. It inhabits a broad range of freshwaters, including Malaysia (Sarawak, Sabah and Peninsular), Thailand,

and Indonesia (Sumatera, Kalimantan, Java). Brunei Darussalam and Laos (Kottelat 2001; Roberts, 1999). Despite its widespread habitat, distant populations of *T. tambroides* may represent different species, which complicates its taxonomy. Until the taxonomy of *T. tambra* has been clarified, it is still categorized as ‘Data Deficient’ by IUCN (Kottelat et al., 2018). It had been reported in previous studies that the appearance of *T. tambroides*, including its coloration (silver, bronze and reddish) varies depending on its locality (Siraj et al., 2007; Walton et al., 2017). This was shown in past research done on portraying the reddish and silver-bronze scales observed were the similar fish species and such attribute can be due to environmental factors with no further solid evidence (Siraj et al., 2007).

Morphologically, *T. tambra* can be differentiated from *T. tambroides* by its lower median lobe and blunt rostral hood (Jaafar et al., 2021) (Figure 1.2). Nonetheless, it shares close similarities with other *Tor* spp., such as the absence of an upper median projection, and a caudal fin of equal length (Walton et al., 2017), making it challenging to distinguish from another *Tor* spp. (Jaafar et al., 2021). In year 2013, researchers highlighted another morphological characteristic in differentiating *T. tambroides* and *T. tambra* through their respective median lobe lengths (Kottelat, 2013).



Figure 1.2: Sketching of Empurau (*Tor tambroides*). Illustration credited to Mr. See Ming Wei

1.4 Taxonomical Status of *Tor*

Morphology characteristics served as a fundamental option for the identification of *Tor* species, including median lobe (Jaafar et al., 2021), mouth structure (Walton et al., 2017) and osteocranium structure (Akmal et al., 2022). The authors proposed a contradictory taxonomic relationship among *Tor tambroides*, *Tor tambra* and *Tor douronensis* based on their respective phylogenetic analysis (Walton et al., 2017; Pinder et al., 2019; Jaafar et al., 2021; Akmal et al., 2022). The data-poor description along with mix of languages in their description among Indonesia *Tor* species had brought difficulty in distinguishing *Tor tambra* and *Tor douronensis* (Walton et al., 2017). Hence, *T. douronensis* found in Peninsular Malaysia was deduced as a synonym of *T. tambra* (Kottelat, 2013). Nonetheless, despite its

ambiguous taxonomy, Indonesian researchers insisted on both species as separate species due to the distinct species differentiation genetically and morphologically (Rahayu & Nugroho 2015). This is further supported by the osteocranium structure analysis of four *Tor* species (*T. tambra*, *T. tambroides*, *T. douronensis*, *T. soro*) with significant differences found in terms of size, number and development of the bone (Akmal et al., 2022). In addition, the appearance of different body shapes of Javan specimens is suggested as intraspecific variation which is insufficient to differentiate the species into separate ones (Roberts, 1993). In addition, there is no current evidence showing the polymorphism of the median lobe and lips resulted from environmental influences as no changes had been observed in the fish from different environments within 66 days (Roberts & Khaironizam, 2008).

In terms of genetic identification, studies had investigated on the phylogenetic relationship of *Tor* spp. through various approaches including 16S rRNA gene, Cytochrome b, Cytochrome c Oxidase Subunit I (*COXI*) gene, mitochondrial DNA, ATPase 6/8 gene, Cytochrome b as well as SNP markers and microsatellites to disentangle its ambivalent taxonomical status (Walton et al., 2017; Jaafar et al., 2021; Lim et al., 2021a). Lim et al. (2021a) had emphasized on the failure of the standard 650 bp *COXI* gene fragment in distinguishing *Tor* species at the species level. The researchers had reported on strongly supported monophyletic clustering of *T. tambra* and *T. tambroides* samples were separated based on sampling locations with Sarawak samples exhibiting paraphyletic relationships (Figure 1.3). In addition, SNPs from 15 mitochondrial genes were documented from the 6 populations based on their taxon sampling to differentiate *Tor* species (Lim et al., 2021a).

Thus, to further conserve *Tor* species, uncovering genomic and transcriptomic data is required for the continuous disentangling and addressing questions of various aspects, such as evolutionary biology, species identification, morphological variations and sequences

related to reproduction, immunity, growth and sex differentiation (Jaafar et al., 2021). However, to date, the whole genome and transcriptome sequences of *Tor* spp. were still not available (as of December 2021) except for studies that had reported on mitogenomic sequences of *Tor* species (Lim et al., 2021a), transcriptome data of *Tor tambra* (Lau et al., 2021b), genomic and gut metagenomic data of *Tor tambroides* (Lau et al., 2022; Lau et al., 2023).

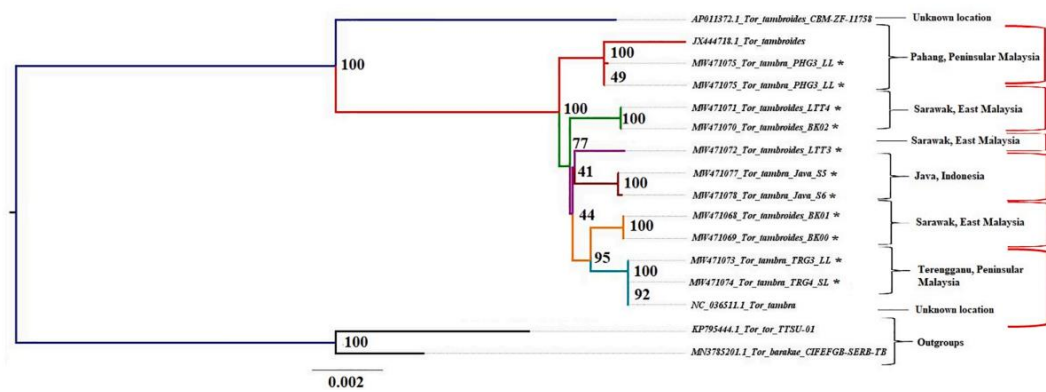


Figure 1.3: Reconstruction of maximum likelihood phylogenetic tree utilizing 13 protein-coding genes and two ribosomal genes across 16 *Tor* species with 1000 bootstraps. Adapted from Lim et al. (2021)

1.5 Aquaculture Diseases

Diseases can be defined as “an unhealthy condition” and “infection with a pathogen” and its occurrence can be due to several reasons: genetic disorders, physical injuries, pathogens, pollution and nutritional imbalance (Kinne, 1980). Fish infection which eventually leads to disease outbreaks are one of the major concerns faced by the aquaculture industry as it would result in mortality and subsequently significant economic damage to the industry. As a common practice, the implementation of high fish-rearing density brought to easier transfer and spreading of pathogens which is believed to be the main cause of such disease outbreaks (Rohani et al., 2022). Stresses exerted due to intensive farming practices

could weaken their innate immune defense and further cause them vulnerable to pathogens (Irshath et al., 2023).

The study of fish diseases concentrates on fish farms where outbreaks can be sudden occurrences with rapid progression and high mortalities and disappear with equal rapidity (acute disease) or slower development with less severity (chronic disease) (Austin et al., 2007). Bacteria-causing diseases do not necessarily caused by single bacteria taxa, instead, there may be a synergistic effect among two or more bacteria taxa. Thus, the occurrence of disease can be the outcome of an interaction between the host (fish), the disease-causing situation (pathogens) and external stressors (unfavorable change in the condition, for example stress). Most of the bacterial pathogens found are one of the members within normal microflora of water and/or water (Austin et al., 2007).

In their natural habitats, fish are described as highly successful due to their resilience to diseases under normal conditions. They are regularly in contact with the environment, possessing high risks to various types of pathogens including both harmless and potential pathogens, which are often in high concentrations (Ellis, 1999; Ellis, 2001; Magnadóttir, 2010). However, the mortalities of fish in the wild could be concealed due to inadequate documentation, behavioral patterns and environmental factors, causing an underestimation of the true impact of the infectious diseases on the fish populations. The reported mass mortalities which further attributed to the occurrence of epizootics or specific disease outbreaks attracted researchers' attention for closer scrutiny to recognize the potential impacts of the diseases as well as to better understand the dynamics of the outbreaks. As in aquaculture industry settings, such an intensive culture system is usually associated with stress which increases the threat of infection as compared to the natural environments.

Moreover, the increased fish density in each tank contributes to the spread of the disease as well (Magnadóttir, 2010).

1.6 *Pseudomonas* Pathogenicity and Prevalence in Environmental and Clinical Settings

Pseudomonas sp. is among the most concerning pathogens for fish and is present in various environments (Pękala-Safińska, 2018). It leads to ulcerative syndrome and hemorrhagic septicemia (Eissa et al., 2010; Algammal et al., 2020). Previous studies have shown that up to half of the farmed fish can be lost to bacterial infection before they are ready for market (Aly, 2013; Algammal et al., 2020). Its pathogenicity has been documented across a wide range of fish species, including *Pseudomonas tructae* in *Oncorhynchus mykiss* (Oh et al., 2019), *Pseudomonas putida* in *Plecoglossus altivelis altivelis* (Wakabayashi et al., 1996), *Oncorhynchus mykiss* (Altinok et al., 2006), *Pseudomonas baetica* in *Dicologlossa cuneata* (Lopez et al., 2012). While *Pseudomonas* species are commonly found in the normal microbiota of fish, they can become opportunistic pathogens and cause infections in the host under stressful conditions (Aly, 2013). Figure 1.4 shows the functional scope and ecological niche of various *Pseudomonas* species, including plant growth promotion, plant disease, biofilm formation, antibiotics resistance, biocontrol, novel secondary metabolites production, bioremediation, saprophytic as well as acting as human host and variable host.

the other hand, early hematopoiesis occurs in the yolk sac, then transferred in the ICM is observed among rainbow trout and killifish (*Pseudepiplatys annulatus*). In rainbow trout and Atlantic salmon (*Salmo salar*), the main immune response relies on the anterior kidney and thymus upon hatching (Razquin et al., 1990). The lymphomyeloid organ development within marine teleost follows the order: kidney, spleen and thymus, with a greater erythropoietic function in larval spleen as compared to lymphopoietic (Nakanishi, 1986; Schröder et al., 1998). As early as 24 hpf (hours post fertilization), cells of myeloid lineage (proerythroblast, myeloblast, myelocyte) can be found within yolk sac. Besides, between 24 and 30 hpf, mature neutrophils are found to move into the blood circulation and its presence was observed within variety of tissues. At the same time, primitive macrophages are developed and invade cephalic mesenchyme to phagocytose (Herbomel et al., 1999; Herbomel et al., 2001). At 34 hpf, erythrocytes and mature neutrophils are observed within the blood system in the connective tissue surrounding the yolk sac. By 28 hpf, alongside young myelocytes and neutrophils, they migrate through the walls of blood vessels. Subsequently, at 72 hpf, mature neutrophils are detected within various tissues (Willett et al., 1999).

The appearance of IgM within fish immune system varies across various species of fish (Magnadóttir et al., 2005). Before the development of the immune system, there has been a maternal transfer of antibodies to eggs and embryos, which had been shown in various species (Dalmo, 2005). The primary goal of such transfer is to offer protection against possible vertical transfer of pathogen and further aid in the early stage of immune development, for instance phagocytosis and the activation of the complement pathway (Magnadóttir et al., 2005). However, this is followed by a significant reduction in maternal IgM concentration during the larval stage as compared to the ovarian stage. There is a 100-fold reduction in its concentration until no maternal IgM detected within sea bass

(*Dicentrarchus labrax*) larvae by 5 dpf (day post fertilization) (Breuil et al., 1997) just before the maturation of larval lymphoid organs and autologous IgM production (Takemura, 1993).

The immunocompetence of fish relies on lymphocytes rather than the histological development of lymphoid organs, and the onset of lymphocyte differentiation varies among different species of fish. Furthermore, despite the early emergence of lymphoid organs and T- and B-lymphocytes, achieving full immunological competence takes a significant amount of time (Zapata et al., 2006). Besides, the development of cell-mediated immunity and T-dependent antibody production precedes the humoral immune response and T-independent antibody production, respectively (Zapata et al., 2006) (Table 1.1).

Table 1.1: Maturation of immunological competence in carps. Adapted from Zapata et al. (2006)

Lymphocytes in Lymphoid organs	Appearance of cells	Ig+	Allograft rejection	Humoral immune response
Thymus (3d -5 m)	Spleen (1m)		16 days	4w (no response to SRBC)
Kidney (6d – 6m)				8w (response to both HGG + FCA and <i>Aeromonas salmonicida</i>)
Spleen (8d – 10m)				

SRBC: sheep red blood cell; HGG: HGG: Hen egg white lysozyme; FCA: Freund’s complete adjuvant

1.8 Fish Immunity

The immune system protects the host from diseases through identification and elimination of pathogens and further suppression of emerging tumour. In addition, it aids in the maintenance of homeostasis during growth and development, especially during tissue damage and inflammation as well (Janeway & Medzhitov, 1998; Du Pasquier, 2001). Classically, immune system is composed of innate and adaptive systems. The innate system, which is evolutionarily ancient and present in all metazoans (Figure 1.5), is encoded within

the germ line and characterized by relatively nonspecific recognition parameters. Thus, the specificity of innate response can be considered an inheritable trait, with the organism's genetic makeup and evolutionary history shaping its response to pathogens, influenced by environmental factors and pathogenic associations (Du Pasquier, 2004; Alvarez-Pellitero, 2008). As illustrated in Figure 1.5, fish are among the earliest animal phyla to develop both innate and adaptive immune systems, making them a compelling model for developmental immune studies (Carroll & Janeway, 1999; Magnadóttir, 2010).

As compared to innate immunity, the emergence of adaptive immunity is considered as a development that occurred relatively recent throughout the evolutionary history, with their first appearance among jaw vertebrates (Gnathostomata). The key element bringing to the evolution of adaptive system is believed to be the rise of thymus, recombination activation (RAG) genes, T- and B-lymphocytes (Du Pasquier 2000; Bowden et al., 2005). Their gene arrangement gives rise to extensive diversity in the immunoglobulin (Ig) superfamily, including major histocompatibility complex (MHC), B- and T-cells receptor (TCR). As compared to inherited innate parameters, these parameters are generated somatically and primarily during ontogeny, resulting in a diversified and limitless pathogen recognition repertoire (Harding & Neefjes, 2005; Huttenhuis et al., 2005). These traits are not inheritable and it reflects the unique immune experience of each individual. In contrast with the instant activation but short duration innate response, the activation of adaptive response is slow due to the selection of specific receptor, cellular proliferation and protein synthesis. However, its effects are enduring (Dixon et al., 1995; Randelli et al., 2008; Magnadóttir, 2010).

It is fascinating that the abrupt emergence of the adaptive immune system coincides with the appearance of Gnathostomata, which can be linked to the primary selection pressure

on jaw development, leading to increased food intake. Consequently, there might be a chance of increased alimentary tract injuries, metabolic rates, mobility and even greater exposure to parasites that further leads to the development of adaptive immunity (Matsunaga, 1998; Rolff, 2007; Magnadóttir, 2010). Later in year 2018, immunologists suggested a possible inheritance of leukocyte populations in vertebrate evolution to retain invertebrate features correlated to the recognition and removal of antigens to some extent. Genes encoding for immunity tends to accumulate in mammals during evolution, forming layered immune system and further into a more complex immune system. As a common thought, “ontogeny resembles phylogeny” suggests the development of an organism reflects its evolutionary history, so a “layered immune system” may provide insights into the development and functionality of immune cells in vertebrates (Scapigliati et al., 2018).

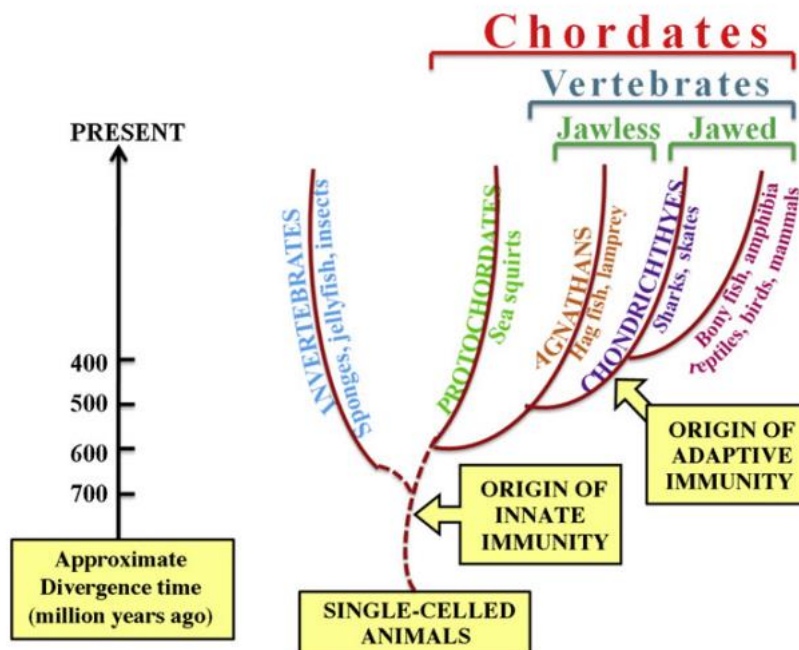


Figure 1.5: The origin and evolutionary connections between innate and adaptive immunity. Adapted from Rauta et al. (2012)

1.8.1 Innate Component

The innate immunity acts as the prime defensive system in fish due to several limitations in fish including slow proliferation, maturation and memory of lymphocytes, limitations in the adaptive system, poikilothermic nature as well as limited antibody repertoire (Whyte, 2007). The system is divided into three components including cellular and humoral components as well as the epithelial/mucosal barrier. It serves as a first defensive barrier in gills, skin and alimentary tract as these are constantly immersed in media containing potential pathogens (Magnadóttir, 2010). Within the fish mucus, a number of immune defense parameters are present, including complement factors, antimicrobial peptides and immunoglobulin, inhibiting the entry of pathogens (Magnadóttir, 2010; Uribe et al., 2011; Rauta et al., 2012).

There are three ways to activate the complement system in teleost: classical pathway (triggered by binding of antibody to cell surface) (Holland & Lambris, 2002), alternative pathway (initiated by foreign microorganisms and do not depending on antibodies), and lectin pathway (activated upon the attachment of mannose-binding lectin to bacteria) (Sakai, 1992). Studies suggest the importance of alternative pathways in the innate response of fish due to its prominent presence within fish serum (Yano, 1996). Complement system is playing essential role in linking both innate and adaptive system in fish despite of killing pathogens through opsonization and phagocytes activation (Carroll, 2004). However, it exhibited bactericidal activity against non-virulent Gram-negative bacteria while showing limited effectiveness against Gram-positive and virulent Gram-negative bacteria (Holland et al., 2002). In addition, studies have shown the importance of antimicrobial peptides in host defence against viruses and bacteria (Ellis, 2001). These molecules are secreted from the skin and can disturb antigen-like bacterial molecules (Lazarovici et al., 1986). Natural

antibodies are found in high levels within serum can offer immediate and broad protection against bacterial and viral invasion. It is linked to adaptive immunity as well (Whyte, 2007).

In fish, the non-specific cellular cytotoxicity is executed by cytotoxic cells which are morphologically distinct from mammals despite their similar functions (Evans & Jaso-Friedmann, 1992). In fish, phagocytosis serves as a crucial defensive mechanism, given its minimal susceptibility to temperature variations owing to the fish's poikilothermic nature (Blazer, 1991). The main macrophages in fish are neutrophils and macrophages with reactive oxygen species production during respiratory bursts for the removal of bacteria (Secombes & Fletcher, 1992). Besides, tumour necrosis factor (TNF) has been suggested as an important activators of macrophages in fish (Uribe et al., 2011). Interferon (INF) are the cytokines that possess non-specific anti-viral function which inhibit nucleic acid replication within the infected cells, protecting the host against viral infection (Robertsen, 2006). In addition, interleukins (IL) in fish are found to be analogous to mammals which play important role in inflammation and host defence as well (Magnadóttir, 2010). Within the body fluids and serum of fish, several protease inhibitors are present to maintain body fluid homeostasis (Bowden et al., 1997). Together with C-reactive protein (CRP) and serum amyloid protein (SAA), they are associated with acute phase response of inflammation (Bayne & Gerwick, 2001; Magnadóttir, 2010). Upon tissue injury, infection or trauma, there will be an elevation in the level of these proteins, with different production pattern between different fish species (Szalai et al., 1994). Transferrin, a protein with a high iron chelator activity, functions by transporting iron ion within fish as iron is crucial in infection establishment by pathogens (Ellis, 2001). Furthermore, lysozyme secreted by monocytes and neutrophils can hydrolyse peptidoglycan of the bacterial cell wall that is found across serum, mucous membrane as

well as kidney and intestine as these tissues are rich in leucocytes (Grinde et al., 1988; Uribe et al., 2011).

Broadly speaking, these innate factors serve as pattern recognition proteins or receptors (PRR), responsible for the identification of pathogen-associated molecular patterns (PAMPs). PAMPs typically consist of diverse polysaccharides and glycoproteins associated with microbes, and they can also detect danger signals emanating from malignant or apoptotic cells (Janeway, 1989; Magnadóttir, 2010).

1.8.2 Adaptive Component

Fish are the most primitive organisms to own adaptive immune systems with Igs as the key humoral parameter. Adaptive immunity will be activated upon the successful invasion of pathogens through the physical barrier and innate immunity (Firdaus-Nawi & Zamri-Saad, 2016). After pathogens invade through the innate immune system, adaptive immunity will recognize and respond to the specific pathogens to avoid the recurrence of infection. The adaptive immune system depends on two major lymphocytes to activate host defence: humoral response (B cells produce Igs) and cellular response (T cells produce lymphocytes) (Randelli et al., 2008).

Upon activation of a specific antigen, either in soluble form or when bound to MHC markers on antigen-presenting cells (APCs), B cells undergo proliferation, giving rise to long-lasting memory cells and plasma cells that secrete specific antibodies. The B1 subset of teleost fish resembles that of mammals in that they produce and release antibodies solely of the IgM class (Wilson et al., 1997; Randelli et al., 2008). Typically, a memory response develops before a second exposure to an antigen (Arkoosh & Kaattari, 1991).

There are differences in the exposure numbers for the fish immunity to respond with two exposures needed by T-dependent antigens whereas once exposure is only needed by the administration of T-independent antigen. In addition, the secondary response is found to be in a faster and larger extent as compared to the primary response. This can be due to the expansion of the memory B cells pool as it had been suggested that the number of antigen-specific B cells within the spleen is directly proportional to the amount of B-cell specific antigen precursor (Kaattari, 1992).

T-cells recognize pathogens associated with MHC markers on their APC through specific markers. Such recognition will produce memory T-cells, continues to persist after infection and will act again when re-encounter (Randelli et al., 2008; Magnadóttir, 2010). T cells can release factors for coordination with other immune cells or cytotoxic factor to kill the infected or abnormal cells (Rauta et al., 2012). There are two categories of T cells found to present within teleost classified according to T-cell receptors (TCR) on the cell surface: alpha-beta ($\alpha\beta$ -) and gamma-delta ($\gamma\delta$ -) TCR. Together with TCR coreceptors, the gene patterns indicated the presence of T cell subpopulations in mammals, including regulatory (Treg, Th17), helper (CD4) and cytotoxic (CD8) (Nakanishi et al., 2011; Scapigliati et al., 2018). Researchers had emphasised the wide distribution of T cells within mucosal tissue of the intestine and gills in fish (Fischer et al., 2006; Buonocore et al., 2006) with the intestine as the possible site for T cells production while gills as the site for T cells to act as effectors and helper (Picchiatti et al., 1997; Scapigliati et al., 2018).

Fish possess all the fundamental components of an adaptive immune system. However, compared to mammals, the diversity of antibodies is relatively limited. This restriction may be attributed to constraints in genetic mechanisms, the lower metabolic rate typical of poikilothermic animals, limited organization of clonal selection, the absence of

restricted presence of lymph nodes or germinal centres, limited memory capabilities, and an absent or restricted capacity for enhanced secondary responses (Du Pasquier, 1982). On the other hand, fish's adaptive immunity system may compensate for such limitations in some aspects as compared to mammals, including a more flexible and loosely associated tetrameric IgM in teleost. Besides, there were three heavy Ig chain classes found in fish: IgD, IgT/Z and IgM with relative higher concentration of IgM present within fish serum as natural antibodies and might have pattern recognition proteins/receptors (Magnadóttir, 2010). IgT/IgZ are monomeric mucosal immunoglobulins (Danilova et al., 2005) with presence of IgT found in trout mucus (Hansen et al., 2005). IgD has been studied at the molecular level with a molecular weight of 150 kDa with little discovery in its physiological role (Ramirez-Gomez et al., 2012).

1.9 Factors Affecting Fish Immune System

1.9.1 Inherent Factors in Shaping Immune Response in Teleost

Differences in disease resistance among various species or individuals can be influenced by inherent factors within the fish immune system, such as evolutionary traits, genetic diversity, age, sex and maternal influences. The evolutionary emergence of teleost, a diverse group of bony fish, can be traced back to the Jurassic period, with distinct modern teleost orders separated by up to 100 to 200 million years.

This group of fish exhibit remarkable adaptability to a diverse range of environments, especially their feeding and breeding habits bringing to more extensive variation. Such diversity suggests the action of selective pressure on teleost's evolution significantly, further influencing their immune system and response. For instance, manifestations are showing

independent evolution involving distinct lines of immune defence mechanisms observed during the comparison of immune response within *Gadidae* (codfish) and *Salmonidae* (salmon) (Pilström et al., 2005). In addition, genetic content also plays a crucial role in fish immunity. Studies have emphasised on the diverse in immune parameters among individuals of same species can reflect on their respective pathogen resistance. Such diversity is able to strengthen the survival of the species under natural conditions (Røed et al., 1993; Van Muiswinkel et al., 1999; Magnadóttir, 2010).

Age and sexual maturation represent crucial inherent factors that significantly impact the immune system and its response in fish. Innate system appears earlier than immune parameters during the ontogeny stage of fish with phagocytes, complement proteins and lysozyme activity as the foundation of the immune system (Magnadóttir et al., 2005). In later stages, as fish reach full development, immune parameters change with increasing age and may exhibit variations between female and male individuals. Studies have observed the continuous increment of Ig present within the serum to rise even throughout life in cod while maximum Ig levels are observed in other species of fish (Klesius, 1990; Nagae et al., 1993). Furthermore, it is clear that the immune response during sexual maturation and spawning can be affected by the interaction of both female and male sex hormones (Maule et al., 1996). The differences in phagocytic and complement activities following the administration of male and female sex hormones were observed within the gilthead sea bream (*Sparus aurata*) (Cuesta et al., 2007).

In aquaculture, the survivability and the fitness of the larvae is usually reflected in the health status of brood fish before and during spawning. Maternal transfer of immune parameters to the larvae can reduce the larvae's vulnerability to potential pathogens after hatching (Mousseau & Fox, 1998). Such transfer may involve both innate and adaptive

components, for instance complement proteins, antimicrobial peptides and lysozyme (Ellingsen et al., 2005). Among them, maternal transfer of immunoglobulin receives much attention, especially IgM to overcome the delayed abilities of larvae to produce antibodies (Takemura & Takano, 1997). Cases had been observed where juvenile sea bream acquire passive protection against specific diseases after receiving maternal antibodies transferred from brood fish vaccinated against the particular disease (Hanif et al., 2005). Commonly, these transferred maternal antibodies existed at low concentration and for a short duration, underscoring their role in providing defence during vertical infection transmission (Swain & Nayak, 2009).

1.9.1 External Factors in Shaping Immune Response in Teleost

External factors can affect fish immunity and response greatly and these factors include temperature as well as stress inducers. As a poikilotherm, environmental changes can significantly affect the fish immune system and such changes can bring to disease prevalence in wild fishes (Zapata et al., 1992; Hayward et al., 2009). In aquaculture industry settings, factors like temperature, light and water salinity are controlled parameters maintained at optimal levels for fish reared in indoor tanks. However, outdoor settings including open-air tanks, earthen tanks or sea cages would raise the environmental fluctuations, leading to the intricate interplay between thermal effects and stress induced by temperature variations. This is further supported by studies which observe the immune response of cod in temperature ranging from 10 °C to 19 °C with minimum expression levels in certain immune parameters including MHC class I, IgM light chain and β_2 -microglobulin at 16 °C (Pérez-Casanova et al., 2008). It is frequently observed that regardless of the influence of stress or pathogens, innate immune parameters tend to exhibit higher activity at

low temperatures, while adaptive immune parameters are suppressed under such conditions (Bly & Clem, 1992). Further studies had demonstrated innate immune activity extends across a broader temperature range, with notably higher activity observed at lower temperatures in its complement-, phagocytic-, natural-cell-mediate-cytotoxicity- (NCC) and lysozyme-activity as compared to endothermic organisms (Lange & Magnadóttir, 2003; Magnadóttir et al., 2005).

Chronic stress is widely acknowledged to tend to suppress the immune response, thereby reducing disease resistance even further. Numerous studies had been conducted to investigate on the impact of stress onto immune system through the monitor of stress hormones, including adrenaline, cortisol and heat shock proteins. In addition, several innate and adaptive parameters together with plasma glucose level are taken into account as well (Gornati et al., 2004; Fast et al., 2008). In relevance to the aquaculture industry, high density, confinement, handling and transport are the main stress inducers which received considerable attention. Studies have reported on the recovery of immune parameters to basal level following short exposure to handling stress in rainbow trout (Demers & Bayne, 1997) and Atlantic salmon (Fast, et al., 2008). Interestingly, fish in the wild appear to show greater stress response to confinement as compared to captive types (Barnett & Pankhurst, 1998). There were various effects found but generally suppressive effects are observed upon direct stress hormone or neuropeptide administration. For instance, down-regulation of Ig-positive lymphocytes was observed after cortisol injection in Atlantic salmon (Espelid et al., 1996) but with stimulation of leukocytes phagocytic activity in the kidney observed within rainbow trout and common carps (Watanuki et al., 2000).

1.10 Pathogen-Host Association during Initial Stage of Disease Manifestation

A pathogen can cause a diverse range of clinical signs and susceptibility towards different fish species and the key to this lies in the difference in the immune defence of each individual. Emphasis had been highlighted on the different histological responses in Atlantic salmon and cod to atypical furunculosis (Wiklund & Dalsgaard, 1998; Magnadóttir et al., 2002). From the difference in histological responses, researchers had concluded that the salmonids tend to show more specific antibody response as compared to cod, showing limited or totally no antibody response (Lund et al., 2008).

It is in good agreement that the transfer of pathogen-host association from one fish species to another is a rare occasion. A delicate balance exists between the mechanisms of infection employed by the pathogen and the defence mechanism of the host. In intensive aquaculture, mortalities may result from a compromised host immune defence or during a newly established pathogen-host association. There are cases when a commensalism or symbiosis relationship cannot be created despite a well-established host-pathogen association, eventually causing the survival of pathogens further for rapid proliferation and even host death (Sitjà-Bobadilla et al., 2006; Magnadóttir, 2010).

Quillet et al. (2007) emphasised the key factor for pathogen infectivity lies on its route of infection which can cause different pathogenicity and severity among different fish species with various host response. Commonly, the primary strategy of pathogens to delay or evade the host's immune response targets the innate immune system. This includes their ability to limit their infections to immune-privileged host tissues (eyes, brain, gonads) where the immune response is naturally suppressed. Antigen-based strategies through antigen mimicry and masking are known in pathogens to inactivate or avoid the humoral defense system (Magnadóttir, 2010). In addition, bacteria tend to produce adhesive molecules or

secret protease to digest mucus for the penetration of the fish integument system (Fang et al., 2004). The protective external capsule is believed to protect the bacteria against the host's humoral parameters and opsonization. Besides, the capsule can hide the virulence factors of the bacteria to be detected by the host and further assist in the spreading of the bacteria. Moreover, some bacteria have developed sophisticated iron-uptake system to counteract the effects of the host's transferrin, securing iron, and generating enzymes to resist the bactericidal activity of phagocytes (Ellis, 1999). The neutralization of respiratory burst product from macrophages appears to be one of the measures for bacteria to endure within phagocytes (Gutenberger et al., 1997).

Studies have emphasized on the activation of the adaptive immune system during the early stage of infection despite the attempt of pathogens to evade the immune response. In many instances, it can effectively eliminate or at least restrict persistent pathogens in the later stage of infections or offer protection upon re-encounter (Raida & Buchmann, 2009). This can be seen in lower bacterial load and the innate response observed when the survivor of the experiment was re-infected with the same bacteria (Raida & Buchmann, 2009). While there are instances where specific antibodies are the primary factors in protecting against bacterial gill disease in rainbow trout, the improved survival rate is generally believed to be due to the adaptive cellular response mediated by specific antigen-stimulated T-cells and cytokine expression (Lumsden et al., 1995; Secombes et al., 1996).

1.11 Antimicrobial Resistance (AMR) and Factors Attributing to it

With the employment of antibiotics for decades for medical uses and preventative measures in various fields, the development of antimicrobial resistance (AMR) is unavoidable (Gajdács & Albericio, 2019; Bhardwaj et al., 2022). The ability of microbes to resist the impact of the antibiotics to which they were previously susceptible, enabling them to survive and develop, is defined as AMR (Zaman et al., 2017). Resistance was first noted in *Staphylococci*, *Streptococci* and *Gonococci*. Soon after the introduction of penicillin, the inaugural commercial antibiotic, in 1942, penicillin-resistant *Staphylococcus aureus* surfaced in 1942 (Dodds, 2017). Likewise, methicillin, a semi-synthetic antibiotic introduced in 1960 as an effort to combat penicillin-resistant *S. aureus*, also encountered resistance in the same year (Durand et al., 2019). AMR has been a significant issue, as the time for bacteria to develop resistance is short, and over 70% of pathogenic bacteria was found to resist at least one type of antibiotic (Economou & Gousia, 2015).

AMR poses significant challenges to public health, with over 2.8 million antibiotic-resistant infections reported annually across the United States. In 2019 alone, these infections resulted in more than 35,000 reported deaths (European Centre for Disease Prevention and Control, 2020). Moreover, a child succumbs to an antibiotic-resistant bacterial infection approximately every 9 minutes. In India, over 50,000 newborns are at risk of death from sepsis caused by antibiotic-resistant microbes (Subramaniam & Girish, 2020). *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae* and *Staphylococcus pneumoniae* were among the most commonly identified antibiotic-resistant bacteria (Uddin et al., 2021). In a recent study, it was found that 3.5% of current TB cases and 18% of previously treated TB cases worldwide developed multidrug-resistant (MDR) or rifampicin-resistant (RR)

tuberculosis (TB) (Migliori et al., 2020). In 2017 alone, there were reported 230,000 deaths worldwide due to MRD- or RR-TB (Ballesterio et al., 2020).

Microorganisms adapt to their surroundings and evolve over time to survive, replicate and spread as swiftly as possible (MacGowan & Macnaughton, 2017). Antimicrobials serve to stop the spreading and growth of bacteria and allow genetic modifications for bacteria to subsequently gain immunity against the medication (Munita & Arias, 2016). However, several factors might still contribute to the complex etiology of antibiotic resistance and are categorised under two main classifications: microbial (natural) factors and antibiotics misuse and overuse.

1.11.1 Microbial (Natural) Factors

AMR can be due to alterations within the bacteria in a variety ways, including genetic mutations, genetic material transfer and selective pressure. Point mutations occurring during bacterial replication may lead to substitution in one or even a few amino acids, subsequently affecting the crucial target, including cell walls, cell structures or enzymes as well as control genes or chromosomal structures (Uddin et al., 2021). As a result, new resistant strains along with their newly acquired immune response could render antibiotics ineffective (Von Wintersdorff et al., 2016).

In addition, genetic material transfer existed within bacteria enable non-resistant bacteria to acquire resistance from drug-resistant bacteria to resist antimicrobials. Antibacterial resistance genes, predominantly found within plasmids and other mobile genetic elements, facilitate their transmission to bacteria of varying species and genera, contributing to widespread resistance (Von Wintersdorff et al., 2016).

Selective pressure denotes the environmental conditions favoring the survival and dissemination of organisms harbouring novel mutations or traits (Holmes et al., 2016). During exposure to antimicrobial agents, the microbes are either eliminated or survive with the development of resistance genes. Surviving resistant microbes will multiply and rapidly dominate the microbial population (Zhao et al., 2019).

1.11.2 Antibiotics Misuse and Overuse

Besides microbial factors, antibiotics misuse and overuse will contribute to the occurrence of AMR as well, including inaccurate diagnosis, irrelevant antibiotics prescriptions, self-medications, insufficient and overuse of antibiotics, poor hospital hygiene management, extensive usage in agriculture and aquaculture as well as lack availability of new antibiotics (Uddin et al., 2021).

During disease diagnosis, healthcare professionals may tend to hinge on unreliable or inaccurate knowledge, leading to precautionous or broad-spectrum antibiotics prescription. As a result, it could exacerbate selective pressure and accelerate antimicrobial resistance (Chokshi et al., 2019; Denning et al., 2017). Furthermore, irrelevant antibiotics prescriptions may occur when the physicians prescribe antibiotics even though they are uncertain whether the infection is caused by bacteria or viruses. Antibiotics can be defined as drugs to fight against bacterial infections but are ineffective against viral infection, and resistance may develop (Marc et al., 2016; Ventola, 2015).

Self-medication with antibiotics without a doctor's prescription is not a rare case across Southeast Asia. Consequently, patients may face the possibility of experiencing adverse reactions to drugs, which can obscure symptoms of underlying illnesses and promote

the development of drug resistance (Nepal & Bhatta, 2018). The incomplete course of antibiotics would lead to the survival of bacteria and further resistance development against that antibiotic. This issue was foreseen by the pioneer of antibiotics, Alexander Fleming and a public caution was issued. Similarly, excessive or inappropriate antibiotic usage could also lead to problems, including bacteria modifications to render the antibiotics ineffective against them (Sreeja et al., 2017; Sulis et al., 2020).

Effective hospital hygiene management is essential due to the constant influx of patients, staff and visitors, each harboring their unique microbiome and hosting bacteria on their clothing and bodies, thereby increasing the risk of colonization. Bacteria could spread easily within this environment and further lead to the emergence and spread of AMR if adequate procedures and protocols for cleanliness maintenance are absent (Almagor et al., 2018). Furthermore, AMR can be worsened by the administration of the antibiotic as growth promoters for animals in agricultural and aquaculture industry. Resistance is still able to develop which can easily proliferate into food chains and further circulate into environment, humans and other living organisms (Sreeja et al., 2017; Manyi-Loh et al., 2018).

The pharmaceutical industry's progress in developing new antibiotics, once effective against bacteria resistant to antibiotics, has markedly slowed down due to technical hurdles, limited understanding, and considerable challenges in addressing bacterial physiology, including the intricate structure of Gram-negative cell walls. Additionally, financial and regulatory hurdles have also contributed to this deceleration. Nonetheless, the inevitable emergence of resistance is still a sad truth despite the introduction of newly discovered antibiotics (Sreeja et al., 2017; Ventola, 2015).

Due to the wide distribution of *Tor* genus across the entire Southeast Asia, it had worsened the taxonomy status. Despite various phylogenetic studies using mitochondrial

DNA, COX1, Cytochrome b, and other genetic markers, there is still lacking of comprehensive genome and transcriptome data for *Tor* species, particularly regarding their conservation status and functional genomics (Walton et al., 2017; Jaafar et al., 2021; Lim et al., 2021a). Further research is needed to uncover its full genome and transcriptome sequencing to clarify its species identification, evolutionary relationships and functional traits relevant to conservation efforts (Jaafar et al., 2021).

In this context, a diseased *Tor tambroides* was obtained from a local aquaculture farm, where the initial hypothesis suggested *Aeromonas* as the causative agents. However, upon isolation and identification, it was discovered to be *P. koreensis*. While *P. koreensis* strains from soil environment are well-studied for their role in plant growth promotion (Scales et al., 2014), there is still limited research on fish-pathogenic strains of *P. koreensis*, particularly in terms of their virulence, antibiotics susceptibility and growth characteristics. Additionally, data on specific virulence factors associated with fish-isolated *P. koreensis* and their potential threat to human health, through fish-borne transmission, remain scarce. A comprehensive study is thus needed to understand on the pathogenicity of these strains in aquaculture settings.

While antibiotics have been the immediate solution to disease outbreaks in aquaculture for decades, their overuse and misuse have led to significant concerns such as antimicrobial resistance (AMR), environmental pollution and health risks through tissue depositions (Uddin et al., 2021; Rohani et al., 2022). Although improvements in water quality can enhance survival rates, these measures are insufficient for long-term disease control (Irsthath et al., 2023). There is a clear gap in the development of sustainable strategies such as enhancing the disease resistance in fish species through genetic approaches. Despite the economic and cultural value of *T. tambroides*, research on disease resistance, especially

towards *P. koreensis*, remains limited. Specifically, there is a lack of studies focusing on the detection and identification of disease-resistant SNP markers in *T. tambroides* that can be applied in marker-assisted selection (MAS) for breeding programs. Thus, the aims of this study are:

- a) firstly, to draft and characterise the genomic data of Empurau and
- b) secondly, to characterise the growth characteristics, virulence and antibiotics susceptibility of *Pseudomonas koreensis* isolated from the local aquaculture farm
- c) finally, to identify and detect candidate disease-resistant SNP markers in resistance to *Pseudomonas koreensis* in *Tor tambroides*

CHAPTER 2

EMPURAU (*T. TAMBROIDES*) GENOME SEQUENCING AND ITS PHYLOGENETICS

2.1 Introduction

Tor tambroides (Bleeker, 1854), a *Cyprinidae* family member, is commonly referred to as Kelah or Empurau in Malaysia and Jurung in Indonesia (Jaafar et al., 2021). It is extensively cultivated and harvested in aquaculture and fisheries, primarily for human consumption (Kottelat et al., 2018). As a true mahseer, *T. tambroides* is typically found in fast-flowing waters with rocky bottoms, alongside *Tor tambra* and *Tor dourenensis*, constituting one of the three *Tor* spp. inhabiting the freshwater bodies of Malaysia, and among the 16 *Tor* spp. found globally (Ng, 2004).

T. tambroides, like other *Tor* spp., is facing a decline in its population size due to environmental degradation in its habitat, which has been exacerbated by human activities such as agricultural practices, logging, and deforestation, leading to a reduction in water flow (Ingram et al., 2005). The consequences of these actions are not limited to a decrease in the population of *T. tambroides* and other *Tor* spp., but also affect the aquatic environment. Furthermore, overfishing using hooks, nets, and dynamite poses a significant threat to the survival of *T. tambroides*, however there is little information regarding the extent of this threat (Kottelat et al., 2018).

T. tambroides is found in various freshwater bodies across Brunei Darussalam, Yunnan China, Indonesia (Kalimantan, Jawa, and Sumatera), Laos, Malaysia (Peninsular, Sabah, and Sarawak), and Thailand (Kottelat, 2018). However, due to its widespread distribution, its taxonomy remains unclear as populations in different regions may differ.

The body scales of *T. tambroides* vary in colour, including silver, bronze, and reddish, depending on the locality.

The taxonomic classification of *Tor* species in Indonesia and Malaysia has come under scrutiny due to notable genetic and/or morphological disparities observed among them. Various studies have attempted to resolve this issue by examining the phylogenetic relationships between these species using different genetic markers such as mitochondrial DNA, COX1 gene, Cytochrome b, ATPase 6/8 gene, 16S rRNA gene, microsatellite, and SNP markers (Walton et al., 2017; Jaafar et al., 2021; Lim et al., 2021a). Nonetheless, comprehensive genome and transcriptome sequences of *Tor* species are still lacking, with limited studies available on their conservation status, gut metagenomics, transcriptomic sequences of *T. tambra* and genomic sequencing of *T. tambroides* (Lau et al., 2021a; Lau et al., 2021b; Lau et al., 2023). Although *T. tambroides* had been identified as tetraploid, with $2n=100$ chromosomes (Donsakul & Magtoon, 2008), further genomic and transcriptomic sequencing of *Tor* species is essential. Such comprehensive sequencing efforts would offer a robust tool for addressing species identification, morphological variations, evolutionary biology, sequences associated with the immune system, sex differentiation, reproduction and growth functions. These insights would contribute significantly to conservation efforts for these species (Jaafar et al., 2021).

Despite the significant growth and expansion of the aquaculture industry, resulting in enhanced fish production compared to traditional fishing methods, certain challenge persists. *T. tambroides* stands out in the aquaculture sector due to its exceptional nutritional value and distinct flesh taste. However, the genetic variability of captive-held fishes decreases, making them more vulnerable to infections. To understand the fish immune system and its response to ecotoxicological chemicals, it is necessary to explore the

underlying mechanisms (Lim et al., 2018; Lim et al., 2021b). Despite the challenges posed by the complexity of the fish immune response and the absence of appropriate reagents for classical immunological assays (Salinas & Magadán, 2017). Additionally, given its slow growth rate, it is crucial to explore the growth-related aspects of its genome to enhance fish farming practices. Omics approaches, including metabolomics, proteomics, transcriptomics and genomics, offer the potential for high-throughput outcomes and the discovery of novel insights.

Despite the recent advancement in long-read sequencing technologies (Heather & Chain, 2016), which have seen increased adoption, studies also utilized a combination of Illumina reads with Nanopore/PacBio reads (Austin et al., 2017; Tan et al., 2018; Lim et al., 2021c; Lim et al., 2022). This amalgamation of short and accurate Illumina reads with longer but less precise Nanopore/PacBio reads has shown promise in improving the completeness of the assembled genome. Thus, the objective is to draft and characterise the genomic dataset of *T. tambroides*. A total of 76,419 predicted non-redundant protein sequences had been identified and annotated to KEGG, GO, and COG databases. Additionally, the immune-related and growth-related genomes as well as the phylogenetic relationship of *T. tambroides* with other Actinopterygii fishes using BUSCO supermatrix data were inferred. These data can be used for improving the conservation alongside immune- and growth-related studies of this fish species.

2.2 Materials and Methods

2.2.1 Ethics Statement

All experiments in this study comply with Animals Research: Reporting In Vivo Experiments (ARRIVE) guidelines and were conducted following the Animal Ethics Committee of Universiti Malaysia Sarawak (UNIMAS/TNC(PI)-04.01/06-09(17)).

2.2.2 Sampling and DNA extraction

An adult male *T. tambroides* (Figure S1) was sampled from a local aquaculture farm, with its locality previously reported in Lau et al. (2021c) and Lim et al. (2021). The fish was euthanized, and 50 mg of its muscle tissue was used for genomic DNA extraction using the DTAB-CTAB DNA extraction kit (GeneReach Biotechnology Corp) following the manufacturer's instructions. The fish had been deposited as voucher specimens (voucher ID: ASD03018) in the fish museum located at the Faculty of Resource Science and Technology, Universiti Malaysia Sarawak.

2.2.3 Species Verification

Cytochrome c oxidase subunit I (COI) gene was amplified, and the PCR product was purified and sent for Sanger sequencing. The sequences were analysed using NCBI BLASTn and found to show 100% similarity with the mitogenome sequences of *T. tambroides* reported by Lim et al. (2021) with the GenBank accession number MW471071.1.

2.2.4 Library Construction and Whole Genome Sequencing

Approximately one µg of gDNA was sheared to 350 bp using a Bioruptor and processed using NEB Ultra Illumina library preparation kit for PCR-free library construction (NEB, Ipswich, MA). The library was quantified using a Qubit (Invitrogen) and sequenced on a NovaSEQ6000 (Illumina, San Diego, CA) with 2 x 150 bp run configuration. For Nanopore sequencing, one µg of unsheared gDNA was used to create a library with flow cell R9.4.1 ligation sequencing kit (LSK109) (Oxford Nanopore, UK) according to the manufacturer's instructions. The sequencing was performed by GeneSeq Sdn Bhd. The library was sequenced on two MinION flowcells with base calling performed on the fast5 files using Guppy version 4.4.1 (Oxford Nanopore) in high accuracy mode.

2.2.5 Sequence Data Processing and Assembly

Illumina reads were trimmed using fastp (Chen et al., 2018), while the Nanopore reads were trimmed with porechop (Wick et al., 2017). A hybrid *de novo* assembly was then performed using Wengan v0.2 (Di Genova et al., 2021) with the trimmed Nanopore reads and paired-end Illumina reads. The genome assembly was subsequently polished with nextpolish (Hu et al., 2020) using the alignment of Illumina reads to the genome. Clean reads were processed with Jellyfish v2.3.0 to obtain a frequency distribution of k-mer, producing kmer frequency distributions of 31-mers (Marçais & Kingsford, 2011). The clean Illumina reads were then aligned to the polished genome using Bowtie2 to assess genome mapping rate (Langmead & Salzberg, 2012). These histograms were then analysed using GenomeScope, which estimates genomic size, repeat content and heterozygosity through a kmer-based statistical approach (Vurture et al., 2017). To evaluate various metrics of the *T. tambroides* genome, QUAST v5.0.2 was used with the parameter of read length 150 bp and

max K-mer coverage 1,000. Completeness of the assembled genome and other publicly available fish genomes was evaluated using BUSCO v5.3.0 based on the single-copy orthologs represented in the actinopterygii_odb10 database (Manni et al., 2021).

2.2.6 Detection of Repetitive Sequences and Prediction of Protein-Coding Genes

The assembled genomes were screened and masked for the repetitive sequences using RepeatModeler2 and RepeatMasker, respectively. Repeat-masked genome assembly reads were aligned to previously generated transcriptomic reads (Lau et al., 2021b) using HiSAT2 (Kim et al., 2015). The resulting transcriptome alignment BAM file and the repeat-masked genome assembly were utilized as the input for BRAKER1 to predict protein-coding genes (Bruna et al., 2021). Furthermore, BRAKER2 was employed for additional annotation via the alignment of vertebrate odb10 conserved proteins (Bruna et al., 2021). Both the BRAKER1 and BRAKER2 annotations were merged and filtered with TSEBRA (Gabriel et al., 2021), resulting in a final set of predicted protein-coding genes.

2.2.7 SSR Analysis

The processed reads of *T. tambroides* genome was analyzed for simple sequence repeats (SSRs) using Kmer-SSR (<https://github.com/ridgelab/Kmer-SSR>) (Pickett et al., 2017). The SSR analysis included dinucleotides, trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides and only SSRs with at least four repeats were considered in this study.

2.2.8 Functional Annotation of Protein-Coding Genes

The protein sequences predicted from the genome were subjected to functional annotation using EggNOG mapper (evolutionary genealogy of genes: Non-supervised Orthologous Groups) with a minimum E-value of 0.001. To facilitate functional annotation, predicted proteins were compared against three public databases, GO (Gene ontology), KEGG (Kyoto Encyclopedia of Genes and Genomes), and COG (the Clusters of Orthologous Groups).

Additionally, predicted protein sequences were mapped against several previously published growth- and immune-related genes based on literature review (Overturf et al., 2010; Palti, 2011; Vagner & Santigosa, 2011; Zhenzhen et al., 2014; Hu et al., 2016; Ma et al., 2016; Chandhini & Rejish, 2019; Lin et al., 2019; Dam et al., 2020; Danzamnn et al., 2016; Guan & Qiu, 2020). 82 growth- and 31 immune-related genes (113 genes in total) were obtained from the NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/>) and used as references. Subsequently, the genes were further filtered based on a stringent E-value cutoff of 10^{-10} .

2.2.9 Ortholog Inference and Phylogenetic Construction

In this study, the assembled genome of *Tor tambroides* and other genome sequences obtained from NCBI (National Center for Biotechnology Information) were summarised in Table 2.1. A total of 52 Actinopterygii species including one Sarcopterygii species as outgroup were included. To identify single-copy orthologous genes across all 52 Actinopterygii species and one Sarcopterygii species, BUSCO v5.3.0 (Benchmarking Universal Single-Copy Orthologs) (Manni et al., 2021) was used with default parameters via Actinopterygii_odb10 database (updated 2021-02-19). For the phylogenetic analysis,

all BUSCOs found in both single-copy and multi-copy for each species were included. Next, BUSCO sequences across all 53 species were aligned MUSCLE (Multiple Sequence Comparison by Log-Expectation) (Edgar, 2004). The gaps and unmatched sites of the resulting multiple sequence alignments (MSAs) were trimmed via trimAl (Capella-Gutiérrez et al., 2009) and subsequently concatenated into a supermatrix. A maximum-likelihood (ML) phylogenetic tree was inferred from the supermatrix using IQ-TREE (Nguyen et al., 2015). The phylogenetic tree generated was analyzed and viewed using Molecular Evolutionary Genetic Analysis (MEGA) (Windows 10 system) v11.0.10 was used to analyze and view the generated phylogenetic tree (Kumar et al., 2018).

The whole genome sequence and raw sequencing reads of *T. tambroides* reported in this study are publicly available under the NCBI Bioproject PRJNA708136. The genome annotation and polished genome assembly generated in this study, were also deposited in Zenodo database (<https://doi.org/10.5281/zenodo.6371903>).

Table 2.1: Available genomes of fish species and their genome statistics used in this study for the Maximum-likelihood (ML; IQ-TREE) tree construction. A total of 50 species with two outgroups were reported and classified into their respective residing environment, including freshwater, marine, or diadromous

Species	Order	Genomic Size (Mb)	BUSCO5 completeness	No. of chromosomes	Accession Numbers
<i>Astyanax mexicanus</i>	Characiformes	1335.2	96.6	25	GCF_000372685.2
<i>Austrofundulus limnaeus</i>	Cyprinodontiformes	866.96	92.1	NA	GCF_001266775.1
<i>Betta splendens</i>	Anabantiformes	441.39	98.3	24	GCF_900634795.3
<i>Carassius auratus</i>	Cypriniformes	1820.6	98.3	50	GCF_003368295.1
<i>Colossoma macropomum</i>	Characiformes	1221.8	98.1	NA	GCF_904425465.1
<i>Coregonus clupeaformis</i>	Salmoniformes	2682.6	97.7	40	GCF_018398675.1
<i>Cyprinodon tularosa</i>	Cyprinodontiformes	1086.8	97.5	NA	GCF_016077235.1
<i>Danio rerio</i>	Cypriniformes	1373.5	95.9	25	GCF_000002035.6
<i>Esox lucius</i>	Esociformes	622.57	97.2	25	GCF_011004845.1
<i>Etheostoma cragini</i>	Perciformes	643.08	94.9	24	GCF_013103735.1
<i>Etheostoma spectabile</i>	Perciformes	854.81	94.9	24	GCF_008692095.1
<i>Haplochromis burtoni</i>	Cichliformes	854.59	97.2	NA	GCF_018398535.1
<i>Lepisosteus oculatus</i>	Lepisosteiformes	945.88	91.5	29	GCF_000242695.1
<i>Melanotaenia boesemani</i>	Atheriniformes	865.59	98.8	24	GCF_017639745.1
<i>Oncorhynchus mykiss</i>	Salmoniformes	2341.7	98.5	32	GCF_013265735.2
<i>Oreochromis niloticus</i>	Cichliformes	1005.7	98.6	23	GCF_001858045.2

Table 2.1: continued

<i>Oryzias latipes</i>	Beloniformes	734.06	96.8	24	GCF_002234675.1
<i>Oryzias melastigma</i>	Beloniformes	778.70	97.2	24	GCF_002922805.2
<i>Paramormyrops kingsleyae</i>	Osteoglossiformes	799.42	96.3	NA	GCF_002872115.1
<i>Poecilia reticulata</i>	Cyprinodontiformes	731.62	96.1	23	GCF_000633615.1
<i>Perca flavescens</i>	Perciformes	877.46	98.6	24	GCF_004354835.1
<i>Perca fluviatilis</i>	Perciformes	951.35	97.5	24	GCF_010015445.1
<i>Pimephales promelas</i>	Cypriniformes	1066.4	96.8	NA	GCF_016745375.1
<i>Poecilia mexicana</i>	Cyprinodontiformes	801.71	96.5	NA	GCF_001443325.1
<i>Polyodon spathula</i>	Acipenseriformes	1542.1	91.9	60	GCF_017654505.1
<i>Pungitius pungitius</i>	Perciformes	466.60	98.2	NA	GCF_902500615.1
<i>Pygocentrus nattereri</i>	Characiformes	1222.1	98.0	30	GCF_015220715.1
<i>Salvelinus namaycush</i>	Salmoniformes	2345.5	94.7	42	GCF_016432855.1
<i>Sander lucioperca</i>	Perciformes	901.24	97.6	24	GCF_008315115.2
<i>Scleropages formosus</i>	Osteoglossiformes	784.56	93.1	25	GCF_900964775.1
<i>Simochromis diagramma</i>	Cichliformes	848.83	96.3	NA	GCF_900408965.1
<i>Sinocyclocheilus rhinoceros</i>	Cyprinoformes	1655.8	38.4	NA	GCF_001515625.1
<i>Sinocyclocheilus anshuiensis</i>	Cyprinoformes	1632.7	36.2	NA	GCF_001515605.1
<i>Sinocyclocheilus grahami</i>	Cyprinoformes	1750.3	49.1	NA	GCF_001515645.1

Table 2.1: continued

	<i>Thalassophryne amazonica</i>	Batrachoidiformes	2446.6	94.4	23	GCF_902500255.1
	<i>Xiphophorus maculatus</i>	Cyprinodontiformes	704.32	97.8	24	GCF_002775205.1
	<i>Xiphophorus couchianus</i>	Cyprinodontiformes	688.54	97.5	24	GCF_001444195.1
	<i>Xiphophorus hellerii</i>	Cyprinodontiformes	732.91	96.4	24	GCF_003331165.1
	<i>Chelmon rostratus</i>	Acanthuriformes	644.21	98.9	24	GCF_017976325.1
Marine	<i>Epinephelus lanceolatus</i>	Perciformes	1087.4	97.1	24	GCF_005281545.1
	<i>Gadus morhua</i>	Gadiformes	669.97	91.8	23	GCF_902167405.1
	<i>Hippoglossus hippoglossus</i>	Pleuronectiformes	596.79	97.6	24	GCF_009819705.1
	<i>Plectropomus leopardus</i>	Perciformes	895.71	92.8	24	GCF_008729295.1
	<i>Thunnus maccoyii</i>	Scombriformes	782.42	99.2	24	GCF_910596095.1
	<i>Toxotes jaculatrix</i>	Carangaria	621.50	98.7	24	GCF_017976425.1
	<i>Latimeria chalumnae</i>	Coelacanthiformes	2860.5	72.7	NA	GCF_000225785.1
	<i>Acanthopagrus latus</i>	Perciformes	685.14	98.7	24	GCF_904848185.1
Diadromous	<i>Alosa sapidissima</i>	Clupeiformes	903.58	95.6	24	GCF_018492685.1
	<i>Kryptolebias marmoratus</i>	Cyprinodontiformes	680.66	97.1	24	GCF_001649575.2
	<i>Lates calcarifer</i>	Carangaria	668.47	94.6	NA	GCF_001640805.1
	<i>Oreochromis aureus</i>	Cichliformes	1005.6	97.7	23	GCF_013358895.1

2.3 Results

2.3.1 *Tor tambroides* Genome Characterisation

The summary of genomic sequencing along with its assembly statistics for *Tor tambroides* were presented in Table 2.2. The total length of contigs reached 1,235,011,685 bp, distributed across 44,726 contigs. The longest contig observed spanned 445,922 bp, with a reported GC content of 36.57%.

Table 2.2: Genomic sequencing and assembly statistics of *T. tambroides*

Content	<i>T. tambroides</i>
Total contig length	1,235,011,685 bp
Number of contigs	44,726
Longest contig length	445,869 bp
Contig N50 length	44,262
Contig GC (%) content	36.57
BUSCO Completeness (Actinopterygii_odb10)	
Actinopterygii_odb10: Complete BUSCOs	87.6% (3192)
Complete and single copy BUSCOs	50.5% (1840)
Complete and duplicated BUSCOs	37.1% (1352)
Fragmented BUSCOs	4.6% (166)
Missing BUSCOs	7.8% (282)

Figure 2.1 illustrates the results of K-mer analysis conducted using Jellyfish software and visualized through GenomeScope (K-mer:31) for the assembled clean reads of *T. tambroides* genome in this study. The analysis revealed a heterozygosity level of 0.194% with no heterozygous peak identified at 23. A homozygous peak depth of 46 was observed, indicating the presence of identical 31-mers on both DNA strands. Additionally, the k-mer based statistical approach indicated that 24.675% of the genome is repetitive, while 75.325%

is unique. By division of K-mer number over the K-mer depth, the predicted genome size of *T. tambroides* were estimated. In this study, the detected K-mer number was 69 Gb, leading to a predicted genome size of 1.23 Gb. Furthermore, the Illumina mapping rate of the current assembled was 86.59%, suggesting the successful assembly of most of the genome. Additionally, Figure 2.2 provides a comprehensive analysis of gene number, gene length, exon length, coding region sequence (CDS) length and protein length of *Tor tambroides* in comparison with closely related fish species in terms of taxonomy.

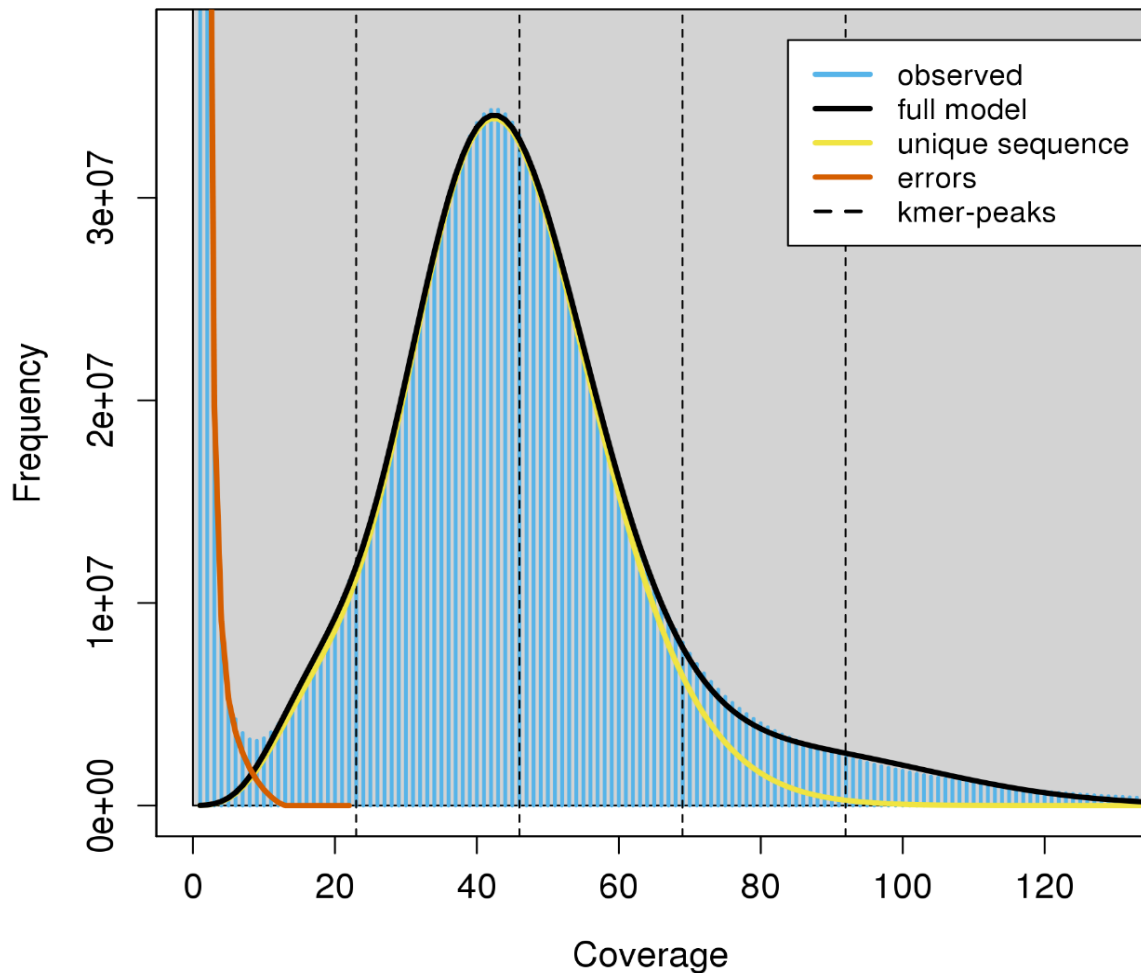


Figure 2.1: Estimation of genome size, repeat content and heterozygosity by GenomeScope, based on 31-mers (read length = 150 bp; kmer max coverage at 1000). The y-axis had demonstrated the amount of K-mers found at each corresponding depth on the x-axis

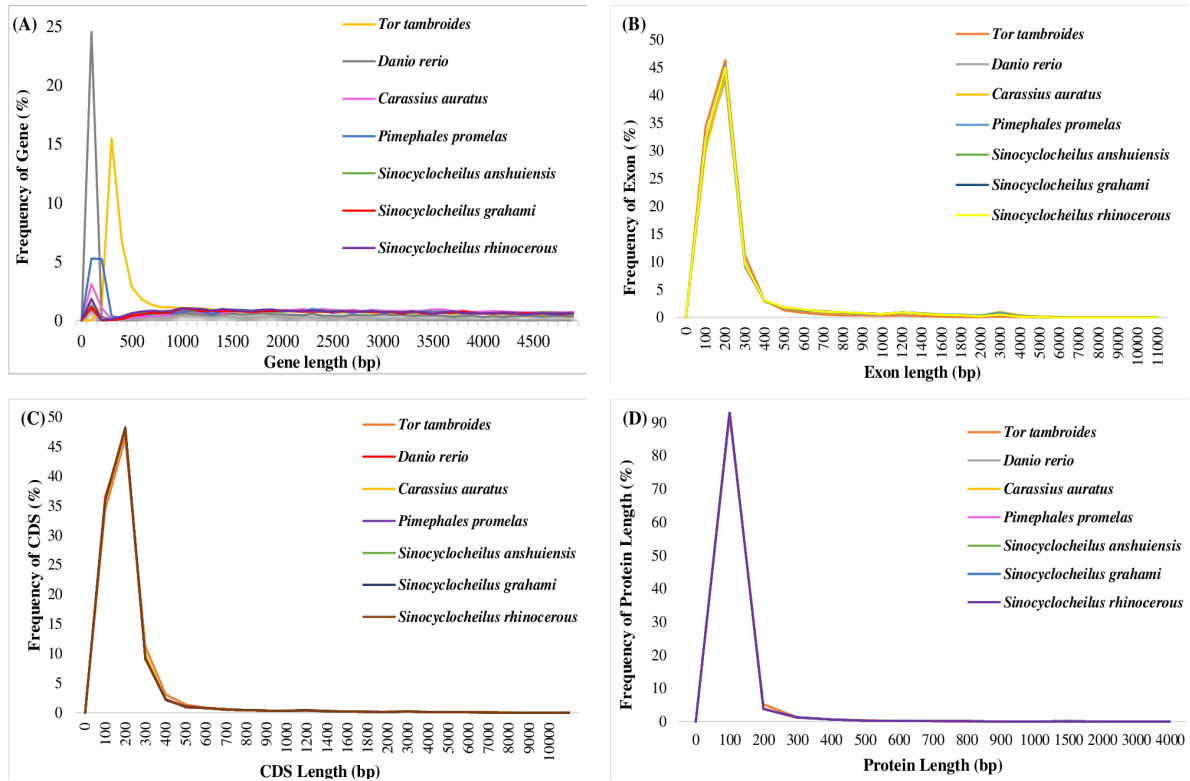


Figure 2.2: Length distribution comparison on total gene, exon, CDS and protein of annotated genes of *Tor tambroides*. Length distribution of gene (A), exon (B), CDS (C) and protein (D) was compared to *Danio rerio*, *Carassius auratus*, *Pimephales promelas*, *Sinocyclocheilus anshuiensis*, *Sinocyclocheilus grahami* and *Sinocyclocheilus rhinoceros*

2.3.2 Repeat-Content Analysis

The analysis of repetitive sequences within the genome of *T. tambroides* via RepeatMasker and RepeatModeler was summarized on Table 2.3. These elements encompass transposable elements (TEs), satellites, simple repeats and sequences of low complexity. A total of 491,441,482 bp of interspersed repeats were documented, consisting of 4,118,572 elements (39.84 %). Among the elements, TEs are classified into two main classes based on the presence of RNA intermediate (class I: retrotransposon) or the absence of RNA intermediate (class II: DNA transposon).

The reverse-transcriptase (RT)-containing retrotransposons, termed long-terminal-repeat (LTR) transposons and non-LTR retrotransposons, can be further classified. LTR

retrotransposon are subdivided into Ty1/copia, Bel, and Ty3/gypsy groups based on the type of RT they possess. Additionally, there is a group known as tyrosine recombinase transposons, which share similarities with LTR retrotransposons but contain integrase instead of recombinase. Non-LTR retrotransposons lack inverted or tandem terminal repeats and feature poly (A) tails at their 3' end, along with variable deletions at their 5' ends encoding open reading frames (ORF), rendering them prone to mutation. Examples of non-LTR retrotransposons include LINEs (long interspersed nucleotide elements), which are autonomous and typically 5-10 kb in length, and SINEs (short interspersed nucleotide elements), which are non-autonomous and approximately 100-400 bp in length (Tang, 2007). Another element constituting only 0.04% of the genome was found as a Penelope-like element (retroelement).

The prevalence of DNA transposons stood out as the highest, comprising 20.33% of the repeats, surpassing unclassified repeats at 10.11%. DNA transposons exhibit distinct traits, featuring short terminal inverted repeats and a lengthy ORF responsible for encoding transposase and DNA binding functions. Employing a 'cut-and-paste' *modus operandi*, this mechanism involves cleavage, replication and relocation facilitated by transposase, observed across numerous eukaryotic organisms (Yuan et al., 2018).

Table 2.3: Summary of repeats within the genome of *T. tambroides*

Type of elements	No. of elements	Length occupied (bp)
Retroelements	192,038	53,452,872
SINEs	17,001	2,346,529
Penelope	2,201	436,250
LINEs	92,074	24,603,740
L2/CR1/Rex	79,288	20,503,943
R1/LOA/Jockey	4,577	972,040
R2/R4/NeSL	158	98,428
RTE/Bov-B	1,011	296,800
L1/CIN4	1,931	847,932
LTR elements	82,963	26,502,603
BEL/Pao	2,371	1,014,621
Tyl/Copia	179	66,269
Gypsy/DIRS1	38,247	14,456,399
Retroviral	23,169	5,730,814
DNA Transposons	1,356,326	251,076,715
hobo-Activator	308,915	53,880,987
Tc1-IS630-Pogo	102,450	36,984,315
PiggyBac	19,086	3,730,730
Tourist/Harbinger	51,882	10,418,360
Other (Mirage, P-element, Transib)	3,510	842,234
Rolling circles	32,781	6,904,609
Unclassified	832,764	124,902,005
Small RNA	12,853	1,250,176
Satellites	122,538	22,386,192
Simple repeats	664,207	28,824,157
Low complexity	74,052	4,026,150

2.3.3 SSR Analysis

Overall, dinucleotide SSRs, notably AT/TA (26.88%), AC/GT (18.19%) and CA/TG (17.78%), dominated the SSR population in the genome of *T. tambroides*, comprising 70% (274, 989 SSRs) of the total 392, 760 SSRs identified. Trinucleotide repeats contributed to 15.30% of the SSRs, while tetranucleotide repeats made up of 12.76%, with the remaining 1.93% attributing to pentanucleotides, summing up to 7,854 SSRs. Figure 2.3 illustrates the three most abundant dinucleotide repeats, AT/TA (105,592), AC/GT (71,461) and CA/TG (69,818). Similarly, the top three dinucleotide repeats were AAT/ATT, TTA/TAA and TAT/ATA, which accounted for one-third of the trinucleotide repeats. The tetranucleotides, AGAT/ATCT, GATA/TATC and TCTA/TCGA were observed with counts of 6,209, 5,386 and 3,826, respectively, collectively representing 30% of all tetranucleotide repeats. The AAAAT/ATTTT (6.32%), ATAAT/ATTAT (6.03%) and TTTTA/TAAAA (5.91%) pentanucleotide repeats had the highest frequencies, covering up to 18% of the pentanucleotide repeat SSRs. More than half of the SSR population within *T. tambroides* genome consisted of dinucleotide SSRs, notably AT/TA (26.88%), AC/GT (18.19%) and CA/TG (17.78%).

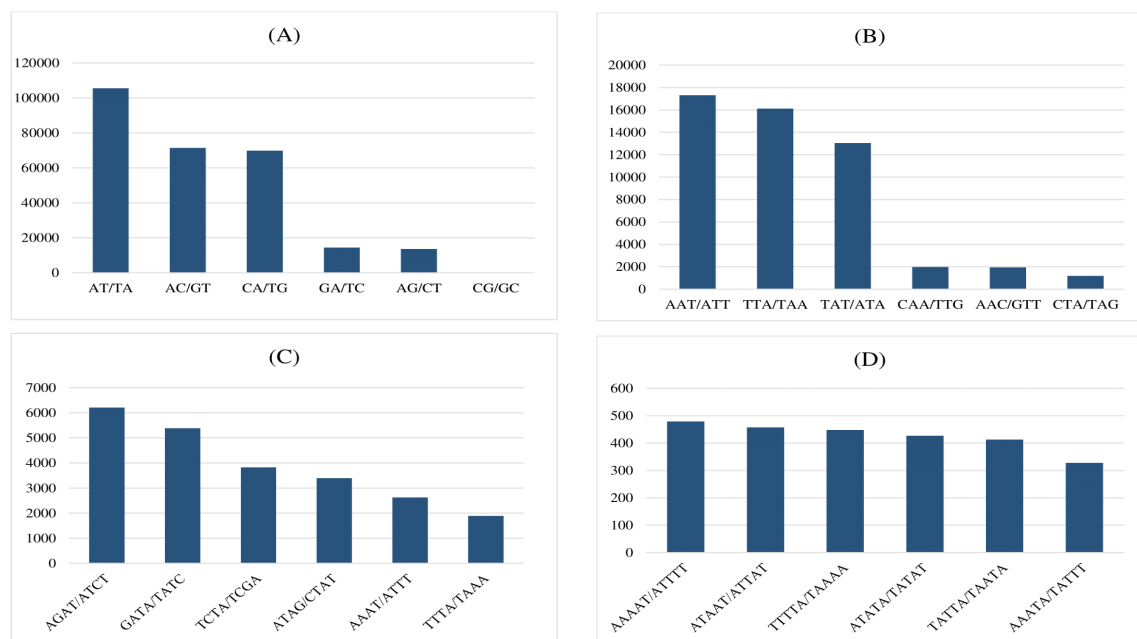


Figure 2.3: SSRs percentage graph with selected top six frequency SSRs from each group: (A) dinucleotide repeats, (B) trinucleotide repeats, (C) tetranucleotide repeats, (D) pentanucleotide repeats

Table 2.4: The number, total length and average length of five different types of SSRs found within the genome of *T. tambroides*

Nucleotide Repeats	Total counts	Total length (bp)	Average length (bp)	Percentage found in the whole genome (%)
Dinucleotide	274989	7935338	28.86	0.022
Trinucleotide	60075	1481739	24.66	0.005
Tetranucleotide	50112	1707712	34.08	0.004
Pentanucleotide	7584	299945	39.55	0.001

2.3.4 Functional Annotation of *T. tambroides*

To annotate *T. tambroides* genome functionally, Interproscan was utilized to extract the coding region (Jones et al., 2014), resulting in 76,419 predicted coding sequences. These sequences were then annotated through EggNOG mapper against GO, KEGG and COG databases with respective number of sequences summarized within Table 2.5. Among them, 20,341 sequences were annotated across all databases. Of the predicted protein sequences, 39,902 (74.48%) were annotated to the GO database, 24,040 (44.87%) to the KEGG database and 53,511 (99.88%) to the COG database. Among the 53,573 annotated predicted protein sequences, 44,519 (83.10%) showed a significant match to at least one database, while 20,342 (37.97%) of the predicted protein exhibited notable matches across all three databases (Figure 2.4).

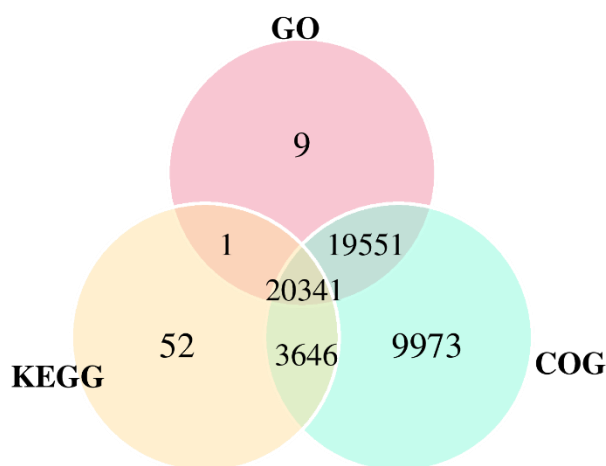


Figure 2.4: Venn diagram showing differences and similarity of predicted protein sequences of *T. tambroides* annotated to GO, KEGG, and COG databases

Table 2.5: Functional annotation of predicted protein sequences to the various database

Database	Number of Predicted Protein Sequences	Percentage (%)
GO	39902	74.48
KEGG	24040	44.87
COG	53511	99.88
Annotated in at least one database	44,519	83.10
Annotated in all database	20,341	37.97

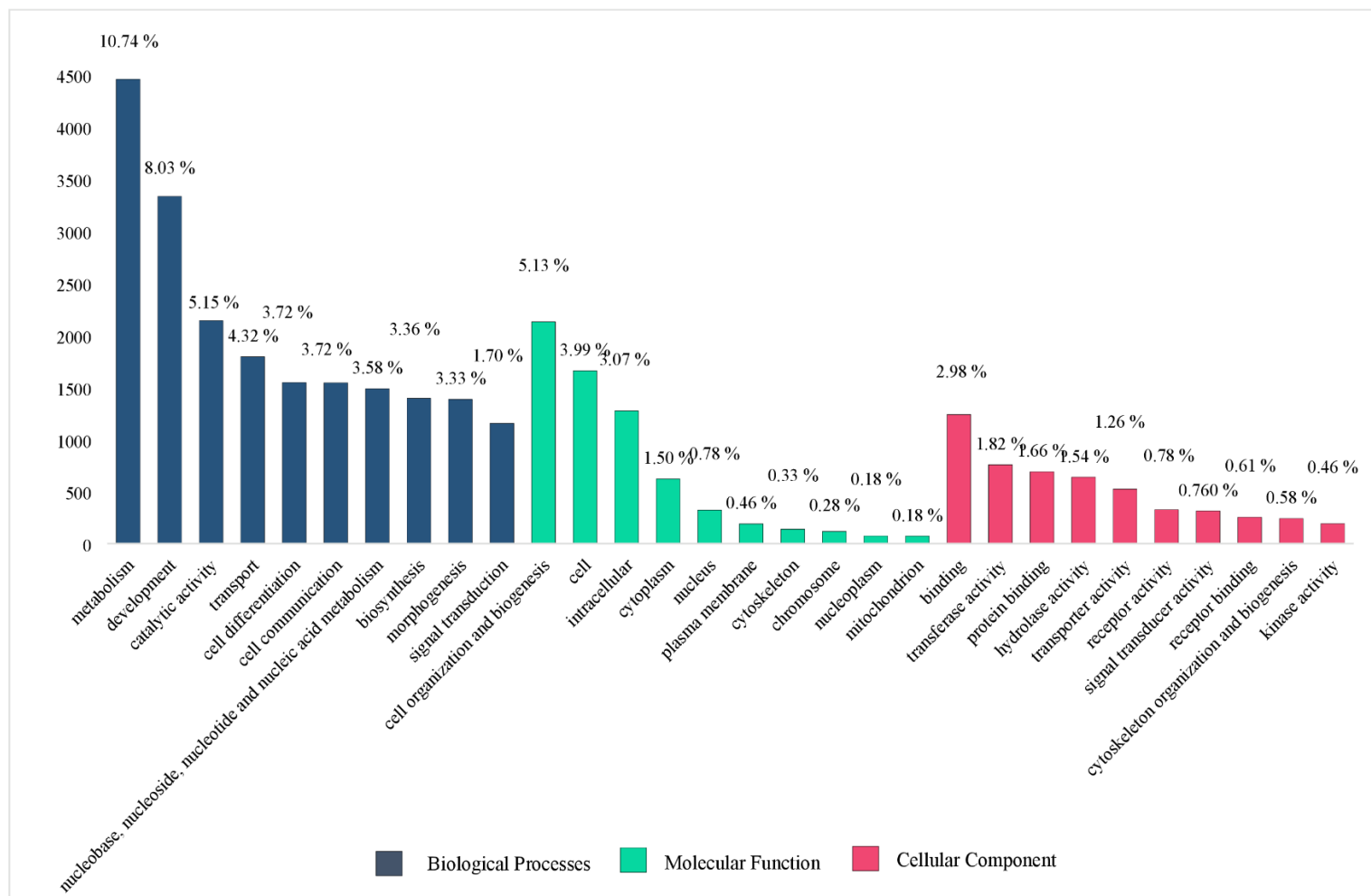


Figure 2.5: GO classifications of *T. tambroides* genome into biological processes, molecular function and cellular component categories

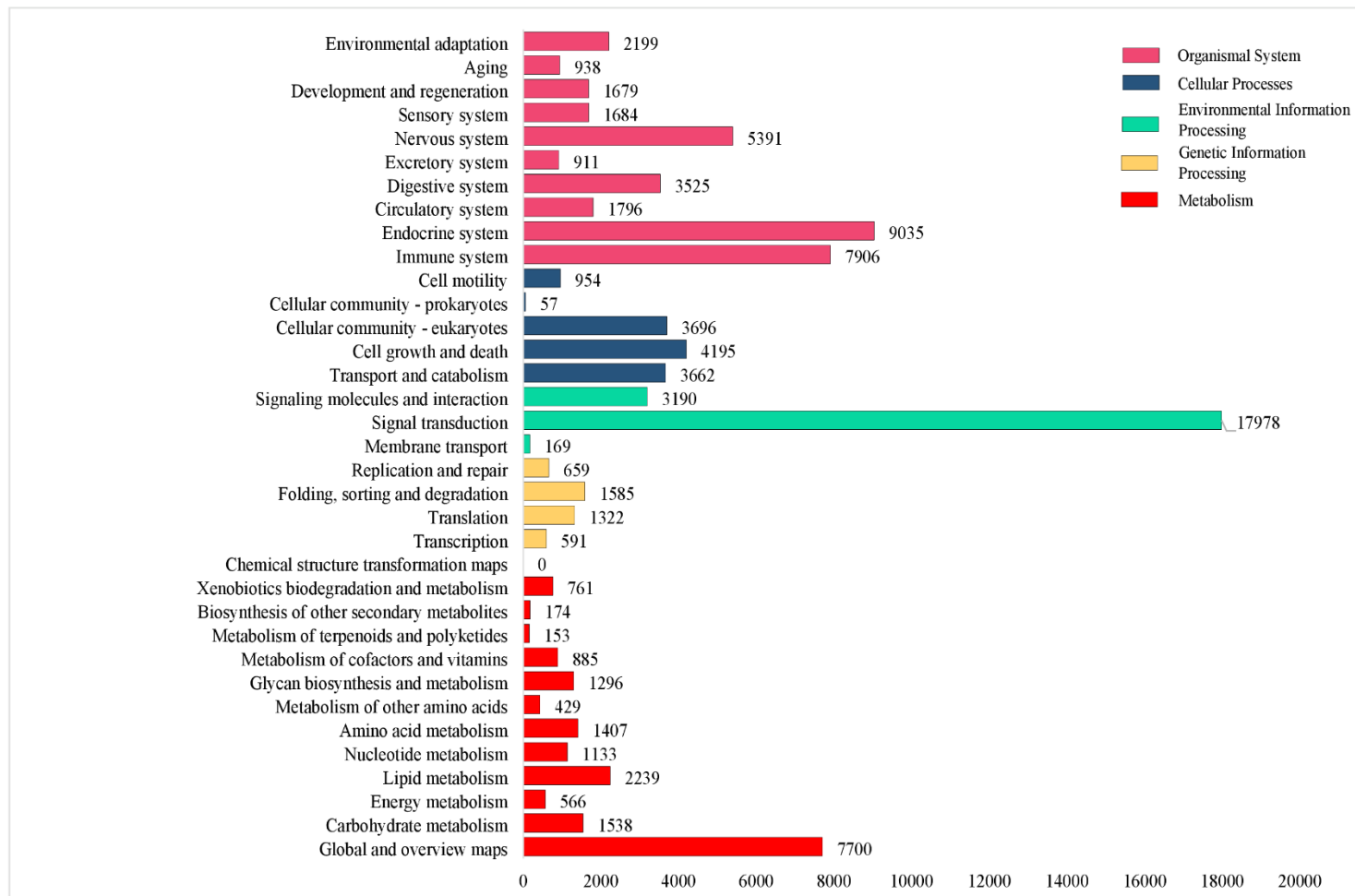


Figure 2.6: KEGG classifications of *T. tambroides* genome into the organismal system, cellular processes, environmental information processing, genetic information processing and metabolism categories

The annotation of *T. tambroides* genome according to the primary ontologies of the GO databases, encompassing biological processes, molecular functions and cellular components was shown in Figure 2.5. Catalytic activity (2141; 5.15%) followed development (3335; 8.03) with metabolism (4460; 10.74%) having the highest count within the biological process category. In contrast, under the molecular function category, binding accounted for a total of 1239 counts (2.98%), while transferase activity and protein binding accounted for 757 (1.82%) and 690 (1.66%) respectively. In the cellular component category, 2130 counts (5.13%) were attributed to cell organization and biogenesis, while 1659 (3.99%) were categorized as cell, and 1274 (3.07%) were labeled as intracellular.

For enhanced integration and interpretation of extensive datasets, the genome underwent additional annotation using the established reference database KEGG, renowned for its diverse pathways. Mapping the genome resulted in successful identification across 304 known KEGG pathways (Figure 2.6), categorized into genetic information processing and metabolism, cellular processes, organismal system and environmental information processing. Genetic information processing had the smallest count (4157; 4.55%), whereas organismal systems boasted the highest count (35694; 38.36%). The categories with the most abundant counts included signal transduction from environmental information processing (17978; 19.67%), endocrine system from organismal system (9035; 9.89%), and immune system (7960; 8.65%) from organismal system. The top ten KEGG cluster components within each of the aforementioned main categories are presented in Figure 2.7. The highest counts were observed in the metabolic pathway (4126; 4.51%), followed by microbial metabolism in diverse environments (1510; 1.65%), and the PI3K-Akt signalling pathway (1406; 1.54%) within the signal transduction category.

The classification of 76,419 predicted protein sequences based on the cluster of orthologous groups in the COG database were presented in Figure 2.8, further grouped into four main clusters: information storage and processing (8971; 15.13%), cellular processes and signaling (24680; 43.08%), metabolism (8023; 14.00%), and poorly characterized (15920; 27.59%). Overall, the genome demonstrates a notable presence of gene families related to signal transduction, the endocrine system, the immune system, and metabolic pathways. This finding aligns with observations from the common carp genome (Xu et al., 2014) and the Javan mahseer transcriptome (Lau et al., 2021b).

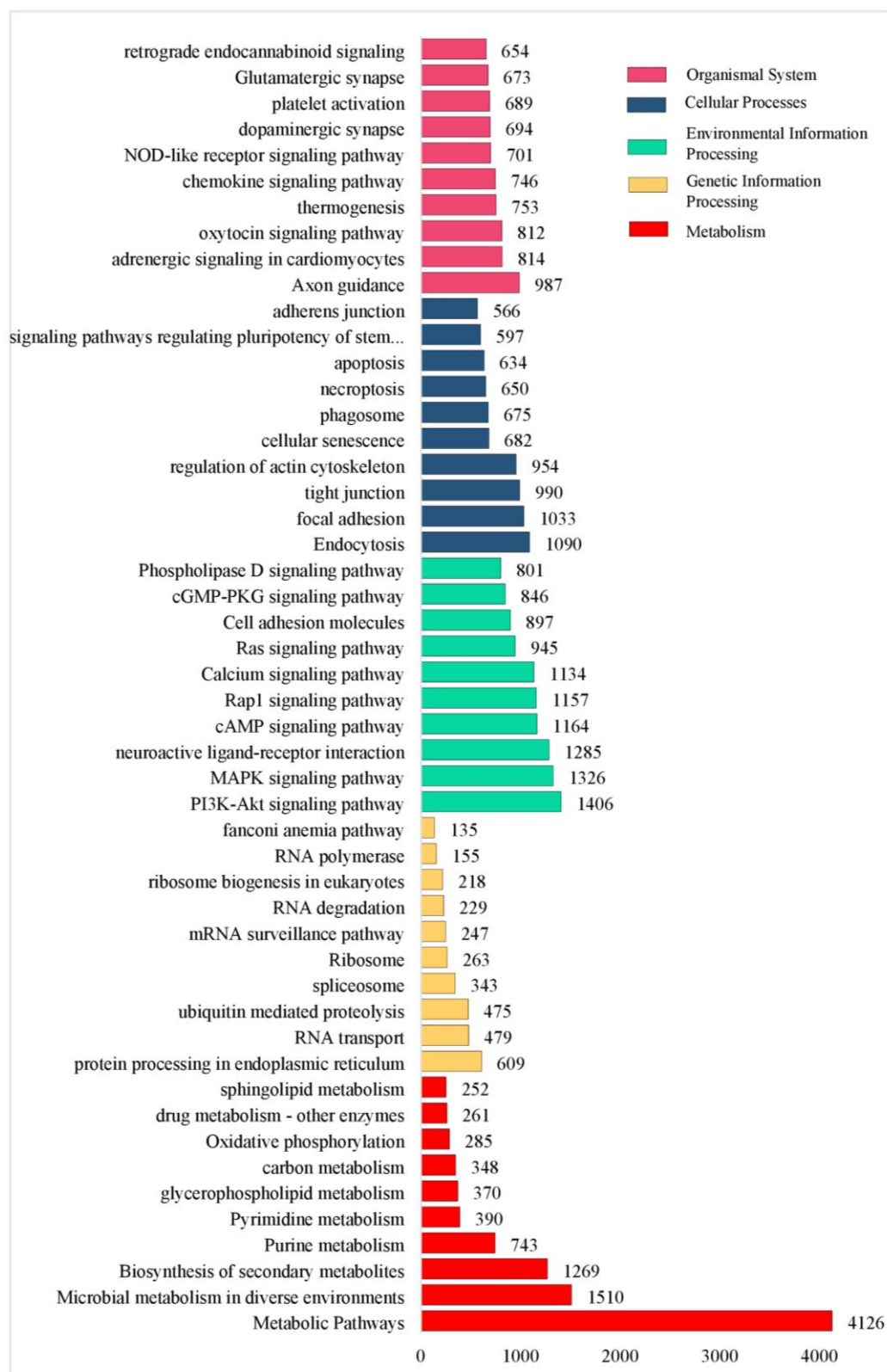


Figure 2.7: Top 10 KEGG classification of *T. tambroides* genome into five main categories: organismal system, cellular processes, environmental information processing, genetic information processing and metabolism categories

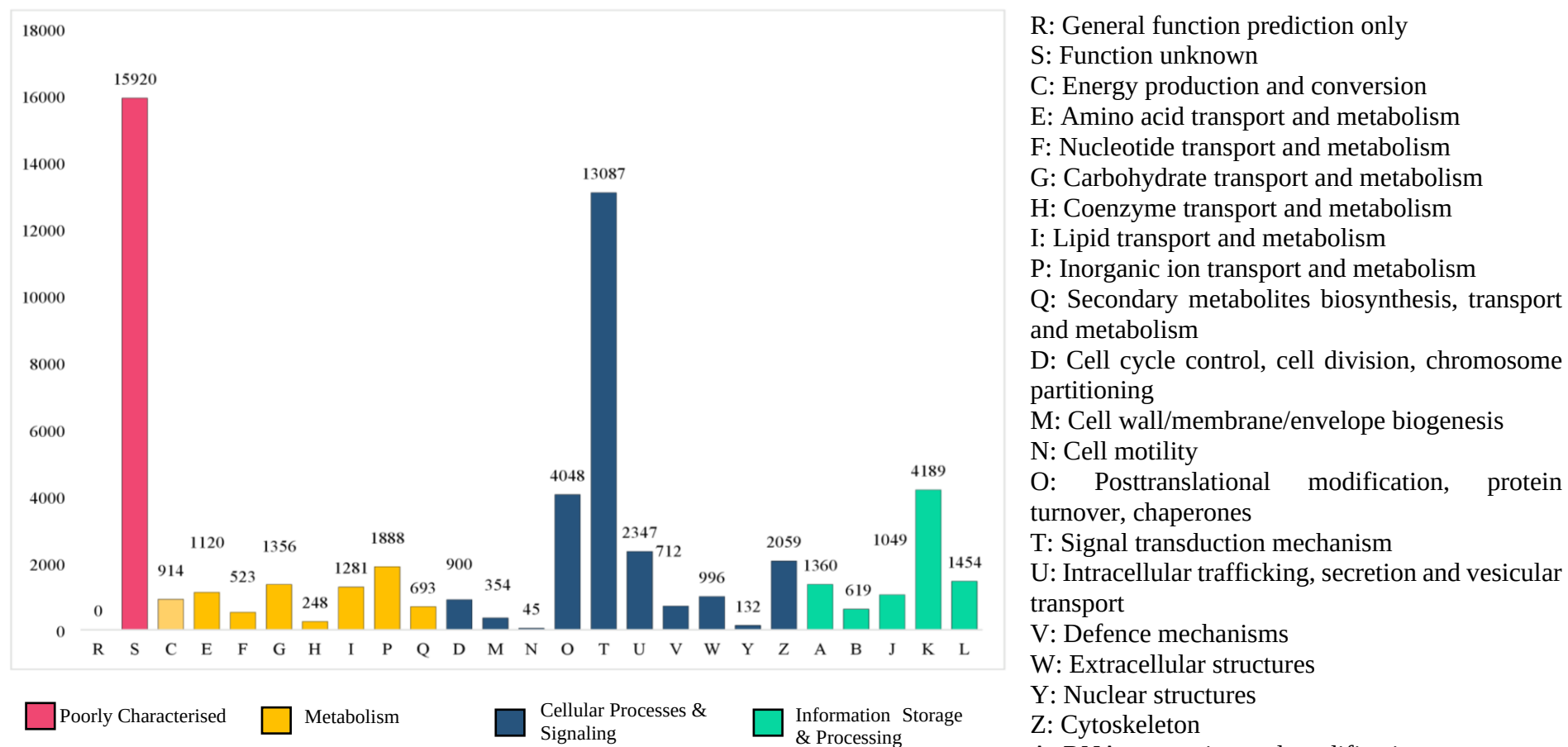


Figure 2.8: COG classifications of *T. tambroides* genome into four main categories: poorly characterized, metabolism, cellular processes & signaling and information storage & processing

2.3.5 Immunity-related Annotation

To gain a comprehensive understanding of *T. tambroides*' defense mechanisms against pathogenic infections, genomic sequences underwent functional annotation to uncover genes and pathways associated with fish immunity. The GO classification revealed that a significant number of the 76,419 predicted protein sequences were assigned immune-related GO terms. Notably, categories such as regulation of the immune system (3926), immune response (3608), and immune system development (2415) showed the highest counts among the annotated categories. KEGG pathway analysis identified 41 immune-related pathways, including the MAPK signaling pathway, Toll-like receptor (TLR) signaling pathway, Wnt signaling pathway, and NOD-like receptor signaling pathway (Figure 2.9) (Zhenzhen et al., 2014). These findings shed light on *T. tambroides*' immune defense mechanisms against pathogenic infections.

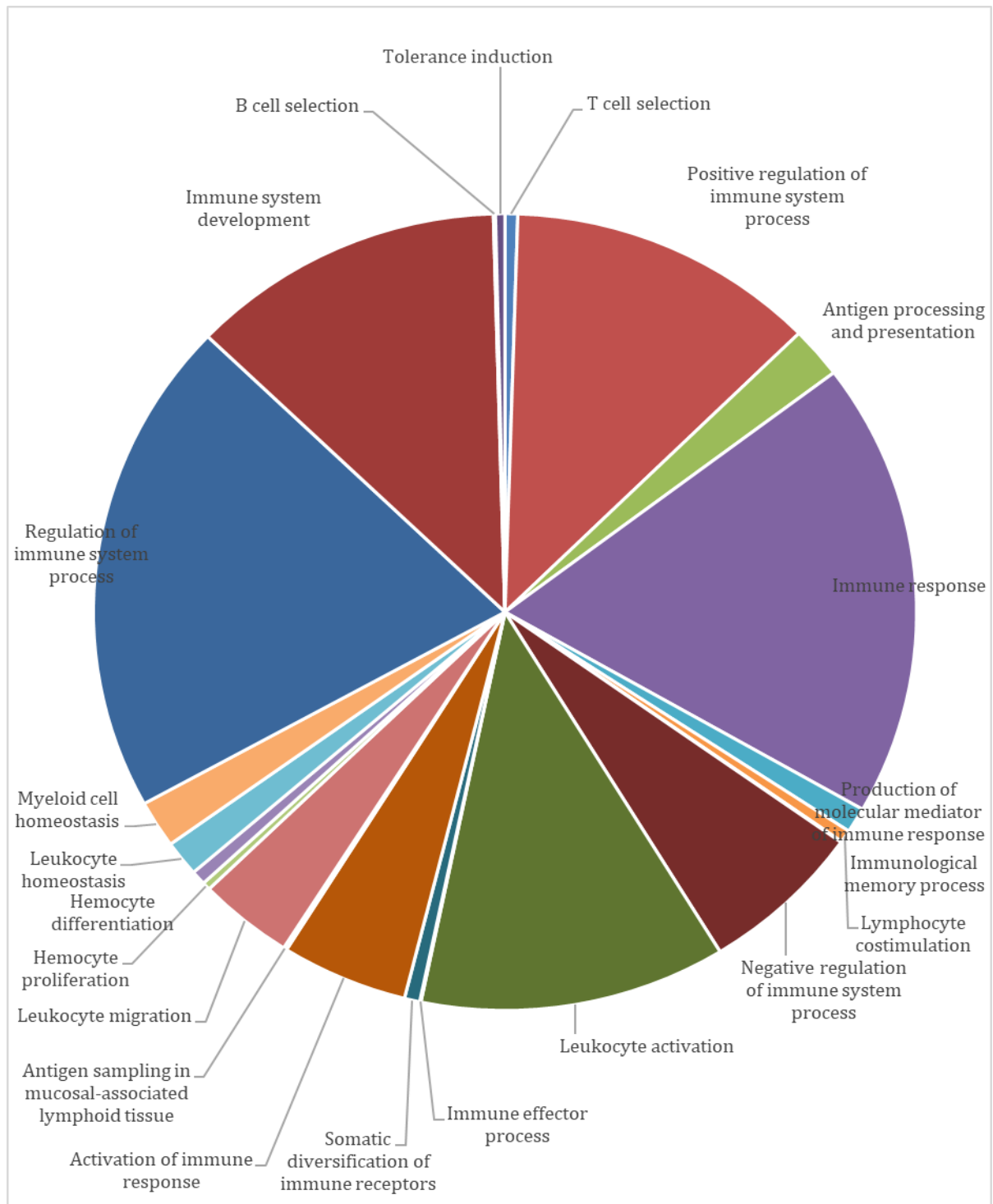


Figure 2.9: Annotation of *T. tambroides* genome to known immune pathway within GO databases

Table 2.6: Immune-related genes. These genes were selected after a stringent e-value cutoff

Gene Location	Gene	Accession	Bit Score	Aligned Length	Similarity (%)	Reference
XP_034741439.1	60S ribosomal protein L8	XP_034741439.1	511	257	98.054	Chandhini & Rejish, 2019
Ttam_g20070.t1	C-C motif chemokine ligand 4	AAT76448.1	99	100	47	Chandhini & Rejish, 2019
Ttam_g8992.t1	C-X-C motif chemokine receptor 4	BAA32797.1	650	338	93.491	Chandhini & Rejish, 2019
Ttam_g53310.t1	Interleukin-10 precursor	NP_001018621.2	293	180	80	Zhenzhen et al., 2014
Ttam_g29226.t1	Interleukin-11	XP_018980434.1	343	190	88.421	Zhenzhen et al., 2014
Ttam_g14739.t1	Interleukin-12 subunit alpha-like	XP_023261134.1	119	194	39.175	Zhenzhen et al., 2014
Ttam_g13738.t1	Interleukin 17a/f3 precursor	NP_001018628.1	181	162	62.963	Zhenzhen et al., 2014
Ttam_g14412.t1	Interleukin 15	NP_001034654.1	231	195	58.974	Zhenzhen et al., 2014
Ttam_g14412.t1	Interleukin 17C	BAV01243.1	231	195	58.974	Zhenzhen et al., 2014
Ttam_g63171.t1	Interleukin 17N	BAV01251.1	253	139	89.209	
Ttam_g46190.t1	Interleukin 21	NP_001122046.1	179	132	69.697	Zhenzhen et al., 2014
Ttam_g47657.t1	Interleukin 22	NP_001018628.1	172	167	53.892	Zhenzhen et al., 2014
Ttam_g20421.t1	Interleukin 34	QCX35396.1	364	209	85.646	Ma et al., 2016
Ttam_g11091.t1	Interleukin 8 precursor	ACI66507.1	137	95	67.368	Zhenzhen et al., 2014
Ttam_g12719.t1	Non classical MHC class I antigen	ABI21842.1	267	338	42.012	Zhenzhen et al., 2014

Table 2.6: continued

Ttam_g370.t1	Myeloid differentiation primary response protein 88	XP_018923074.1	544	284	92.254	Zhenzhen et al., 2014
Ttam_g22246.t2	Toll-like receptor 1	NP_001124065.1	1226	798	75.94	Ma et al., 2016
Ttam_g29226.t1	Toll-like receptor 11	XP_018980434.1	343	190	88.421	Zhenzhen et al., 2014
Ttam_g14739.t1	Toll-like receptor 12	XP_023261134.1	119	194	39.175	Zhenzhen et al., 2014
Ttam_g19676.t2	Toll-like receptor 13	NP_001133860.1	790	925	45.946	Zhenzhen et al., 2014
Ttam_g23835.t2	Toll-like receptor 14	AXL48518.1	973	814	58.845	Zhenzhen et al., 2014
Ttam_g49992.t1	Toll-like receptor 2 precursor	NP_997977.1	1273	787	79.161	Zhenzhen et al., 2014
Ttam_g14350.t3	Toll-like receptor 21	AVX48323.1	1684	959	87.904	Ma et al., 2016
Ttam_g19676.t2	Toll-like receptor 22 precursor	NP_001117884.1	857	954	47.17	Ma et al., 2016
Ttam_g47317.t1	Toll-like receptor 3	ABL11473.1	1604	904	88.496	Palti, 2011
Ttam_g47734.t1	Toll-like receptor 4ba	AHH85806.1	359	212	83.491	Palti, 2011
Ttam_g47734.t1	Toll-like receptor 4bb	AHH85807.1	370	244	76.23	Palti, 2011
Ttam_g36189.t1	Toll-like receptor 5b precursor	NP_001124067.2	1433	878	79.613	Zhenzhen et al., 2014
Ttam_g12988.t1	Toll-like receptor 7	AIS23537.1	1394	1052	67.586	Zhenzhen et al., 2014
Ttam_g18125.t2	Toll-like receptor 9	ADE20130.1	2030	1064	95.019	Chandhini & Rejish, 2019

2.3.6 Growth-related Annotation

To delve into growth-related aspects of the *T. tambroides* genome, functional annotation was conducted using previously identified growth-related genes as a reference. Table 2.7 presented the BLAST results of growth-related genes within the genome, indicating that 82 genes passed the E-value cutoff filter of 10^{-10} . A total of 19 growth-related pathways were revealed via KEGG pathway analysis, encompassing pathways in cancer, the insulin signaling pathway, endocytosis, focal adhesion, and the mTOR signaling pathway (Zhenzhen et al., 2014) (Figure 2.10).

Previously published growth-related genes exhibited significant similarity with the genome of *T. tambroides*. For instance, a high degree of similarity was detected between growth hormone (GH) and insulin-like growth factor-binding protein (IGFBP), which play key roles in regulating GH and IGF levels (Zhenzhen et al., 2014), as indicated in Table 2.7. Furthermore, somatostatins were identified to inhibit GH release (Li & Lin, 2010).

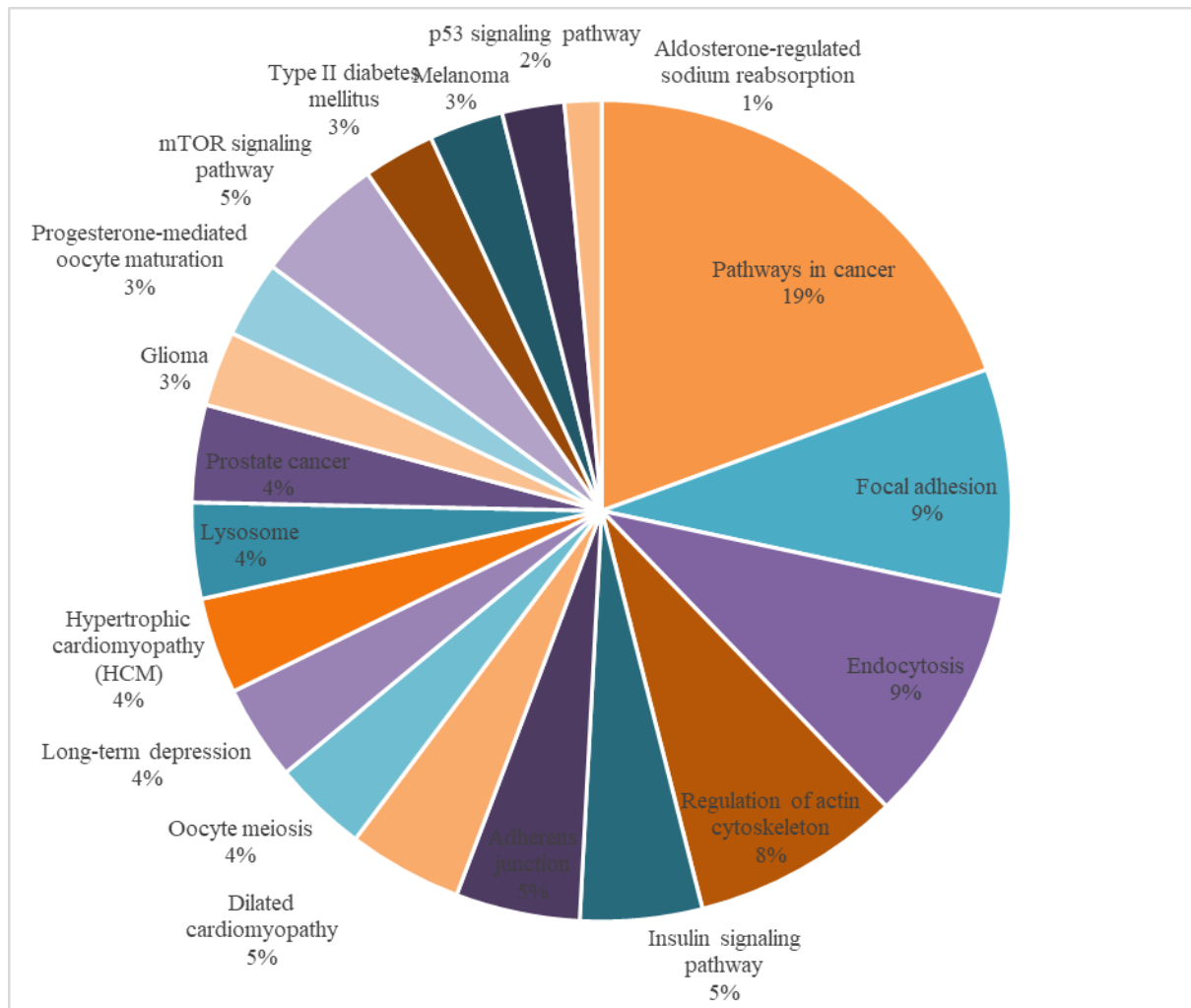


Figure 2.10: Annotation of *T. tambroides* genome to known top 20 growth-related pathway in KEGG database

Table 2.7: Growth-related genes. These genes were selected after a stringent e-value cutoff

Gene Location	Gene	Accession	Bit Score	Aligned Length	Similarity (%)	Reference
Ttam_g16164.T1	acetyl-coa carboxylase alpha (haplotype 1)	ADT82650.1	3652	2370	73.966	Guan & Qiu, 2020
Ttam_g16164.T1	acetyl-coa carboxylase alpha (haplotype 2)	ADX43925.1	3651	2351	74.351	Guan & Qiu, 2020
Ttam_g16164.T1	acetyl-coa carboxylase 2	XP_018962132.1	3753	1837	97.877	Guan & Qiu, 2020
Ttam_g455.t1	acetyl-coa acetyltransferase 2	AAD34966.1	751	395	92.152	Hu et al., 2016
Ttam_g34749.t1	alpha actinin 3	XP_031588180.1	1577	897	82.386	Damzamn et al., 2020
Ttam_g3317.t2	adiponectin	NP_001373470.1	432	254	81.496	Hu et al., 2016
Ttam_g1647.t1	alpha-1-microglobulin/bikunin precursor	XP_005451429.1	189	198	44.444	Dam et al., 2020
Ttam_g36407.t1	apolipoprotein b	NP_001013332.1	407	259	78.764	Dam et al., 2020
Ttam_g10197.T1	atp synthase	BAA82837.1	1035	518	99.421	Dam et al., 2020
Ttam_g6111.t1	atpase h ⁺ transporting v1 subunit e1	AAM34666.1	440	226	96.46	Guan & Qiu, 2020
Ttam_g20808.t1	g2/mitotic-specific cyclin-b1	NP_571588.1	790	398	94.472	Guan & Qiu, 2020
Ttam_g15943.t1	cell division cycle protein 20	NP_998245.2	917	496	90.524	Guan & Qiu, 2020
Ttam_g9535.t1	conserved edge expressed	ABK35126.2	552	322	90.994	Overturf et al., 2010
Ttam_g38788.t1	chymotrypsin-like elastase family member 1	XP_018937349.1	463	264	93.939	Dam et al., 2020
Ttam_g17232.t1	Trypsin inhibitor citi-1 like	XP_017329258.1	88.2	73	53.425	Dam et al., 2020
Ttam_g9508.t1	Creatine kinase m-type	XP_028660838.1	665	381	84.514	Damzamn et al., 2020
Ttam_g47430.t1	Peroxisomal carnitine O-octanoyltransferase	XP_018953213.1	1228	612	94.935	Lin et al., 2019
Ttam_g15021.t1	Cathepsin K precursor	NP_001017778.1	636	333	89.489	Dam et al., 2020
Ttam_g58501.t1	Cathepsin L	AAI08032.1	568	336	76.786	Dam et al., 2020

Table 2.7: continued

Ttam_g38141.t1	25-hydroxycholesterol_7-alpha-hydroxylase	XP_699028.2	878	506	82.016	Lin et al., 2019
Ttam_g6091.t1	Delta-6_desaturase	AZL94116.1	926	444	100	Vagner & Santiagosa, 2011
Ttam_g18373.t1	Elongation of very long chain fatty acids protein 6	XP_018957993.1	438	266	82.707	Dam et al., 2020
Ttam_g47213.t1	Fatty acid binding protein	ACT83304.1	104	111	53.153	Hu et al., 2016
Ttam_g18259.t1	Fatty acid-binding protein 10-A	NP_694492.1_	240	126	90.476	Hu et al., 2016
Ttam_g14995.t2	Fatty acid synthase	ARO92273.1	4851	2517	94.398	Lin et al., 2019
Ttam_g7992.t1	Farnesyl pyrophosphate synthase isoform X1	XP_005472704.1	608	359	81.058	Lin et al., 2019
Ttam_g58962.t1	Forkhead box protein K1 isoform X1	XP_025764132.1	793	687	64.92	Lin et al., 2019
Ttam_g66177.t2	G0/G1 switch protein 2	AVM4087.1	214	112	91.964	Damzamn et al., 2020
Ttam_g18731.t1	Glucose-6-phosphatase	AVP32214.1	655	355	94.93	Dam et al., 2020
Ttam_g47478.t1	Growth hormone	CAA43007.1	385	210	95.238	Ma et al., 2016
Ttam_g50496.t1	Growth hormone receptor	ADZ13484.1	1109	574	94.077	Ma et al., 2016
Ttam_g15275.t1	Growth hormone receptor 1	AAY86769.1	493	616	47.727	Hu et al., 2016
Ttam_g49456.t1	Growth hormone releasing hormone	BBA21042.1	249	141	92.199	Ma et al., 2016
Ttam_g67728.t1	Glutathione synthetase	XP_018970229.1	884	475	90.105	Lin et al., 2019
Ttam_g14998.t1	70-kDa heat shock proteins	AAF70445.1	1128	634	86.278	Hu et al., 2016
Ttam_g28364.t1	Insulin-like growth factor-binding protein 1	ACV72066.1	524	262	96.947	Guan & Qiu, 2020

Table 2.7: continued

Ttam_g46996.t1	Insulin-like growth factor-binding protein 2	ACM47497.1	548	274	97.08	Guan & Qiu, 2020
Ttam_g2941.t1	Insulin-like growth factor-binding protein 3	ACM47527.1	179	271	40.221	Guan & Qiu, 2020
Ttam_g3167.t1	Inositol Monophosphatase 1	XP_018975133.1	571	282	97.872	Lin et al., 2019
Ttam_g31807.t1	Insulin-induced_gene_1_protein	NP_956163.1	503	251	97.61	Hu et al., 2016
Ttam_g49453.t1	Lipopolysaccharide binding protein	NP_001118057.1	737	469	74.414	Hu et al., 2016
Ttam_g22739.t1	Leptin A2	AGK24956.1	309	171	88.889	Zhenzhen et al., 2014
Ttam_g64331.t1	Leptin B1	QEN91938.1	337	168	95.238	Zhenzhen et al., 2014
Ttam_g4584.t1	Tyrosine-protein kinase JAK2-like	XP_022620925.1	1601	967	78.904	Zhenzhen et al., 2014
Ttam_g15162.t1	Hepatic triacylglycerol lipase	XP_018956861.1	964	498	92.169	Dam et al., 2020
Ttam_g28008.t1	Myocyte Enhancer Factor 2	BAA33567.1	381	475	51.579	Overturf et al., 2010
Ttam_g7423.t1	Myostatin	ACS92640.1	513	374	65.775	Overturf et al., 2010
Ttam_g7423.t1	Myostatin 1	AJF48833.1	515	351	68.661	Overturf et al., 2010
Ttam_g7423.t1	Myostatin 2	AJF48834.1	726	366	94.536	Overturf et al., 2010
Ttam_g9041.t1/Tt am_g3292.t1	Myogenic factor 5	AWG41988.1	486	240	97.083	Ma et al., 2016
Ttam_g14955.t1	Neuropeptide Y receptor Y7	NP_001007219.1	714	372	95.161	Zhenzhen et al., 2014
Ttam_g16503.t1	Cytochrome P450	AAK37960.1	709	505	66.535	Hu et al., 2016
Ttam_g58558.t1	Paired box protein 7 isoform X1	XP_013988550.1	555	305	89.18	Overturf et al., 2010
Ttam_g4591.t1	Cytosolic phospholipase A2 gamma-like	XP_018952444.1	976	518	89.189	Dam et al., 2020
Ttam_g36121.t1	Proopiomelanocortin	AAM93491.2	387	222	83.333	Zhenzhen et al., 2014

Table 2.7: continued

Ttam_g66767.t1	Proopiomelanocortin b precursor	NP_001076520.1	278	216	77.778	Zhenzhen et al., 2014
Ttam_g61812.t1	Peroxisome proliferator-activated receptor alpha	CAJ76702.1	699	468	73.077	Hu et al., 2016
Ttam_g26783.t1	Peroxisome proliferator-activated receptor beta	ACR15760.1	690	430	76.279	Hu et al., 2016
Ttam_g15650.t1	Preprosomatostatin I	AJB44559.1	218	114	96.491	Zhenzhen et al., 2014
Ttam_g25535.t1	Preprosomatostatin II	AJB44560.1	210	110	96.364	Zhenzhen et al., 2014
Ttam_g23085.t1	Preprosomatostatin III	AJB44561.1	221	120	95.833	Zhenzhen et al., 2014
Ttam_g61734.t1	Prolactin	CAA31060.1	382	210	95.714	Ma et al., 2016
Ttam_g19109.t1	Prolactin receptor	QIB98245.1	1089	611	86.579	Ma et al., 2016
Ttam_g59684.t1	60S ribosomal protein L10	XP_018921966.1	436	215	97.209	Danzamn et al., 2020
Ttam_g10921.t1	Antithrombin-III	XP_018920986.1	844	450	92	Dam et al 2020
Ttam_g27833.t1/ Ttam_g52890.t1	SMAD family member 3	ABI94729.1	877	423	99.527	Lin et al., 2019
Ttam_g446.t1	Secreted Protein Acidic And Cysteine Rich	XP_003447656.1	519	300	83.667	Lin et al., 2019
Ttam_g17820.t1	Squalene monooxygenase	NP_001103509.1	1024	557	89.408	Ma et al., 2016
Ttam_g43467.t1/ Ttam_g46793.t1	Somatostatin Receptor type 1-like	XP_018943223.1	749	367	98.365	Ma et al., 2016
Ttam_g20275.t1	Somatostatin Receptor type 2-like	XP_018946514.1	603	337	88.131	Zhenzhen et al., 2014
Ttam_g10831.t1	Somatostatin Receptor type 3-like	XP_018963641.1	515	261	98.467	Zhenzhen et al., 2014
Ttam_g51397.t1	Somatostatin Receptor type 5-like	XP_018961210.1	661	443	81.49	Zhenzhen et al., 2014

Table 2.7: continued

Ttam_g2866.t1	Signal transducer and activator of transcription 1b	NP_956385.2	1219	722	80.471	Zhenzhen et al., 2014
Ttam_g43797.t1	Signal transducer and activator of transcription 2	NP_001258730.1	1360	852	78.638	Zhenzhen et al., 2014
Ttam_g12941.t1	Signal transducer and activator of transcription 3	BAH47263.1	1675	806	99.256	Zhenzhen et al., 2014
Ttam_g13118.t1	Signal transducer and activator of transcription 4	NP_001004510.1	1328	679	95.876	Zhenzhen et al., 2014
Ttam_g26382.t1	Signal transducer and activator of transcription 5	BAH47264.1	1587	785	97.58	Zhenzhen et al., 2014
Ttam_g31273.t1	TFAR family member associated NF-kappa-B activator	XP_035460435.1	251	354	44.633	Damzamn et al., 2020
Ttam_g1720.t2	Troponin C	XP_018973573.1	312	160	100	Damzamn et al., 2020
Ttam_g39818.t1	Troponin T	XP_029379803.1	284	216	85.185	Damzamn et al., 2020
Ttam_g47967.t1	Ubiquitin carboxyl-terminal hydrolase 38	XP_003449754.1	1421	1039	70.356	Lin et al., 2019

2.3.7 Orthologs and Phylogenetic Inferences

Using the BUSCO supermatrix approach via single-copy orthologs, the phylogenetic relationship between *T. tambroides* and other ray-finned fishes was established. Table 2.8 summarizes and assesses the BUSCO completeness of each species. Genomic sequences of 52 Actinopterygii species and one Sarcopterygii species as an outgroup were aligned using MUSCLE (Edgar, 2004). Following that, a phylogenetic tree was built utilizing the maximum-likelihood model (ML; IQ-TREE) and relying on both single-copy and multi-copy orthologs (Figure 2.11). With the exception of the outgroup species (Coelacanthiformes and Acipenseriformes), all others belonged to the class Actinopterygii, but they are categorized into 18 distinct orders, which comprise Anabantiformes, Atheriniformes, Batrachoidiformes, Clupeiformes, Cichliformes, Cypriniformes, Characiformes, Cyprinodontiformes, Esociformes, Gadiformes, Perciformes, Salmoniformes, Pleuronectiformes, Lepisosteiformes, Beloniformes, Osteoglossiformes. *T. tambroides*, belonging to the family *Cyprinidae*, formed a monophyletic cluster with species in the order Cypriniformes, including species from the genus *Sinocyclocheilus* as well as *Carassius auratus*, *Pimephales promelas*, and *Danio rerio*.

Table 2.8: BUSCO completeness of 52 ray-finned fishes for the inference of phylogenetic relationship with *T. tambroides* using BUSCO supermatrix approach

Species		BUSCO5 Completeness (%)				
		Complete	Complete and single-copy	Complete and duplicated-copy	Fragmented	Missing
Freshwater	<i>Archocentrus centrarchus</i>	98.2	97.2	1.0	0.4	1.4
	<i>Astyanax mexicanus</i>	96.6	95.5	1.1	1.5	1.9
	<i>Austrofundulus limnaeus</i>	92.1	91.6	0.5	3.6	4.3
	<i>Betta splendens</i>	98.3	97.2	1.1	0.6	1.1
	<i>Carassius auratus</i>	98.3	37.3	61.0	0.7	1.0
	<i>Colossoma macropomum</i>	98.1	97.1	1.0	0.9	1.0
	<i>Coregonus clupeaformis</i>	97.7	51.2	46.5	1.3	1.0
	<i>Cyprinodon tularosa</i>	97.5	96.7	0.8	0.9	1.6
	<i>Danio rerio</i>	95.9	78.5	17.4	1.7	2.4
	<i>Esox lucius</i>	97.2	95.8	1.4	1.2	1.6
	<i>Etheostoma cragini</i>	94.9	94.1	0.8	1.5	3.6
	<i>Etheostoma spectabile</i>	94.9	91.3	3.6	0.8	4.3
	<i>Haplochromis burtoni</i>	97.2	96.2	1.0	0.9	1.9
	<i>Lepisosteus oculatus</i>	91.5	90.2	1.3	3.9	4.6
	<i>Melanotaenia boesemani</i>	98.8	97.8	1.0	0.4	0.8
	<i>Oncorhynchus mykiss</i>	98.5	55.3	43.2	0.7	0.8

Table 2.8: continued

<i>Oreochromis niloticus</i>	98.6	97.8	0.8	0.5	0.9
<i>Oryzias latipes</i>	96.8	96.2	0.6	0.5	2.7
<i>Oryzias melastigma</i>	97.2	96.5	0.7	1.1	1.7
<i>Paramormyrops kingsleyae</i>	96.3	93.5	2.8	1.5	2.2
<i>Poecilia reticulata</i>	96.1	95.5	0.6	1.5	2.4
<i>Perca flavescens</i>	98.6	97.6	1.0	0.4	1.0
<i>Perca fluviatilis</i>	97.5	96.8	0.7	0.8	1.7
<i>Pimephales promelas</i>	96.8	94.3	2.5	1.4	1.8
<i>Poecilia mexicana</i>	96.5	93.5	3.0	1.7	1.8
<i>Polyodon spathula</i>	91.9	51.6	40.3	2.6	5.5
<i>Pungitius pungitius</i>	98.2	97.0	1.2	0.4	1.4
<i>Pygocentrus nattereri</i>	98.0	97.0	1.0	1.0	1.0
<i>Salvelinus namaycush</i>	94.7	54.6	40.1	1.3	4.0
<i>Sander lucioperca</i>	98.6	97.6	1.0	0.5	0.9
<i>Scleropages formosus</i>	96.5	93.1	3.4	1.3	2.2
<i>Simochromis diagramma</i>	97.8	96.3	1.5	0.5	1.7
<i>Sinocyclocheilus rhinoceros</i>	96.9	38.4	58.5	1.3	1.8
<i>Sinocyclocheilus anshuiensis</i>	97.9	36.2	61.7	1.0	1.1
<i>Sinocyclocheilus grahami</i>	93.7	49.1	44.6	2.9	3.4
<i>Thalassophryne amazonica</i>	95.6	94.4	1.2	1.1	3.3

Table 2.8: continued

	<i>Xiphophorus maculatus</i>	98.5	97.8	0.7	0.4	1.1
	<i>Xiphophorus couchianus</i>	98.2	97.5	0.7	0.5	1.3
	<i>Xiphophorus hellerii</i>	98.4	96.4	2.0	0.5	1.1
Marine	<i>Chelmon rostratus</i>	98.9	98.1	0.8	0.4	0.7
	<i>Epinephelus lanceolatus</i>	98.3	97.1	1.2	0.8	0.9
	<i>Gadus morhua</i>	92.7	91.8	0.9	1.8	5.5
	<i>Hippoglossus hippoglossus</i>	98.3	97.6	0.7	0.4	1.3
	<i>Plectropomus leopardus</i>	94.7	92.8	1.9	1.8	3.5
	<i>Thunnus maccoyii</i>	99.2	98.6	0.6	0.4	0.4
	<i>Toxotes jaculatrix</i>	98.7	98.1	0.6	0.3	1.0
	<i>Latimeria chalumnae</i>	75.3	72.7	2.6	5.8	18.9
Diadromous	<i>Acanthopagrus latus</i>	98.7	98.0	0.7	0.3	1.0
	<i>Alosa sapidissima</i>	95.6	94.2	1.4	1.7	2.7
	<i>Kryptolebias marmoratus</i>	97.7	97.1	0.6	0.7	1.6
	<i>Lates calcarifer</i>	98.5	94.6	3.9	0.6	0.9
	<i>Oreochromis aureus</i>	98.5	97.7	0.8	0.5	1.0

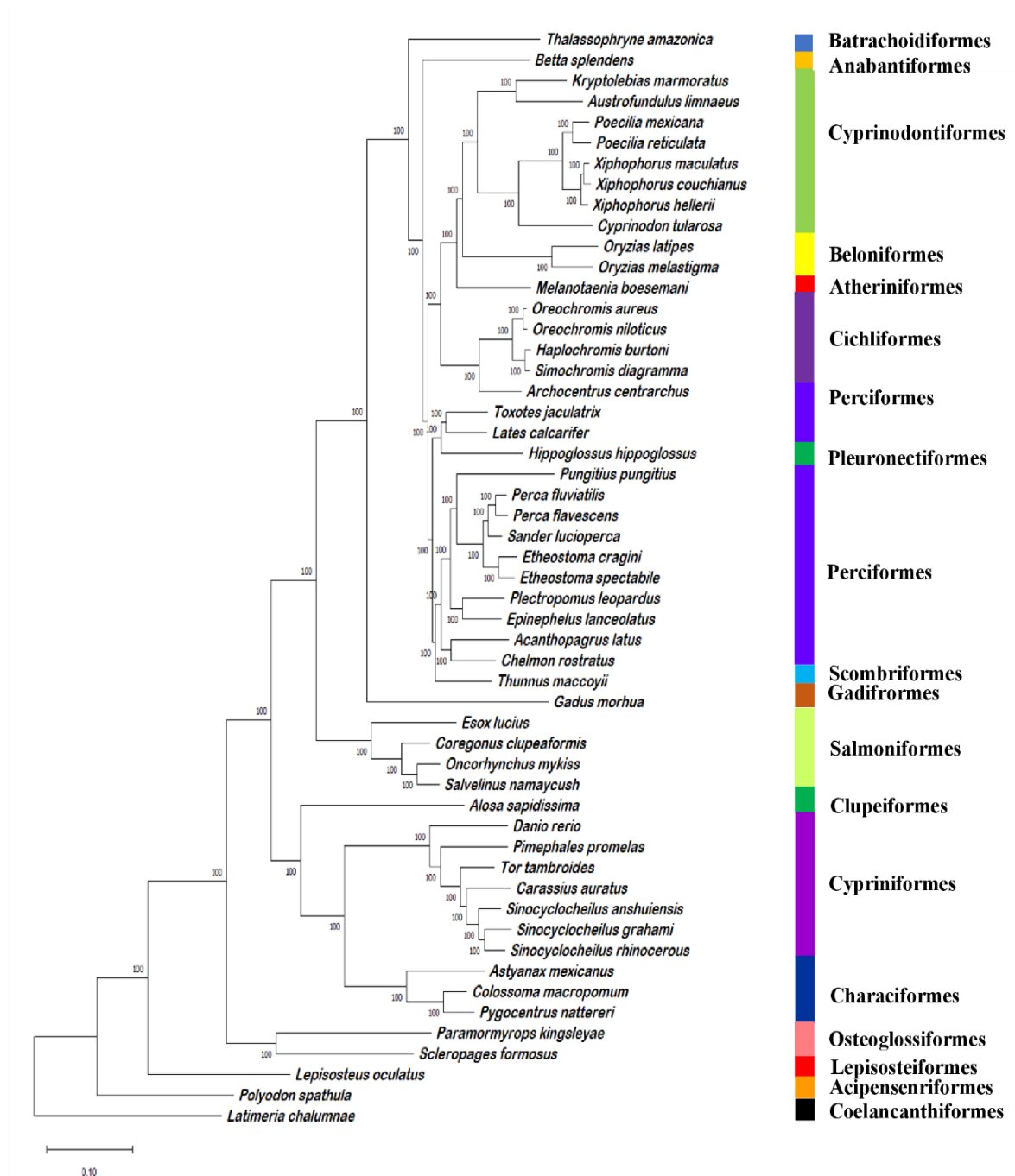


Figure 2.11: Maximum-likelihood tree plotted on 52 Actinopterygii species with one Sarcopterygii species as outgroup inferred from a supermatrix of 3,640 BUSCOs. The species were displayed as order

2.4 Discussion

The discovery of the genomic data is essential for the progression of aquaculture research and hatchery technology, especially in the molecular breeding of empurau. In previous studies, *T. tambroides* was reported to have a chromosome number of $2n=100$ (Donsakul & Magtoon, 2008). This study reveals that the genome size of *T. tambroides* is 1.23 Gb, which is smaller than that of other tetraploid *Cyprinidae* species within the same family. For instance, goldfish (*Carassius auratus*) has a genome size of 1.85 Gb (Chen et al., 2019b), and common carp (*Cyprinus carpio*) has a genome size of 1.83 Gb (Xu et al., 2014). Whole-genome duplication events in tetraploid may result in a higher retention of duplicated genes (Xu et al., 2014).

In addition, the GC% content of *T. tambroides* genome was reported to be 36.55%, slightly lower than the values reported for *C. idella* and *C. carpio* (37.4% and 37.3%, respectively). Typically, seawater fish exhibit a higher GC content, often surpassing 40%, whereas freshwater fish tend to have a GC content below 40%. Additionally, migratory species generally have a higher GC content compared to non-migratory ones (Lu & Luo, 2020). Consequently, it is plausible that various environmental factors influence the genomic GC content (Lu & Luo, 2020).

Intriguingly, the genome of *T. tambroides* exhibited a contrasting pattern between genomic size (1.23 Gb) and GC content (36.55%), diverging from the findings of a previous study that examined 14 species. Lu & Luo (2014) emphasized a converse relationship between GC content and genome size. Thus, genomic data involving diverse fish species was recommended for further validation of such postulation (Lu & Luo, 2020).

2.4.1 SSR analysis

During genome evolution, SSRs are utilised to increase DNA quantity and raw materials for faster adaptation to environmental stress. Therefore, as a matter of fact, despite the conventional expectation that shorter repeats like mono- and dinucleotide repeats are more stable, the mutation rate of microsatellites is influenced by the length of the repeated unit (Schlötterer, 2000). The genome showed the emergence of an unexpected pattern where the frequency of tetranucleotide repeats exceeded that of trinucleotide repeats, echoing findings in other ray-finned fishes (Lei et al., 2021). The unexpected prevalence may be elucidated by considering the role of the triplet code in gene formation and the presence of a mismatch repair system in the exonic region. This repair system likely maintains fewer trinucleotide repeats, consequently leading to a reduction in their number compared to tetranucleotide repeats (Lei et al., 2021). Furthermore, the stability of repeat units was directly perpendicular to the length of the repeat units, suggesting greater structural stability in tetranucleotides as compared to trinucleotides (Sup Lee et al., 1999).

A greater prevalence of SSR repeats containing poly (A/T) tracts compared to those with poly (G/C) tracts across a variety of SSRs, such as dinucleotide, trinucleotide, tetranucleotide, and pentanucleotide repeats were reported by Lei et al. (2021). The higher occurrence of poly (A) can be attributed to the reintegration of mRNA-processed genes into the genome, typically accompanied by a poly (A) tail. Conversely, poly (G/C) does not participate in this reintegration process (Lei et al., 2021). Furthermore, the higher prevalence of poly (A) can be elucidated by the formation of pseudogenes and their involvement in the universal retrotransposon process (Tóth et al., 2000; Borodulina et al., 2016). In contrast, (GC) repeats exhibit greater stability compared to (AT) repeats and are less prone to slippage

during replication. As a result, GC-rich sequences are commonly found in coding regions across both prokaryotes and vertebrates (Oliver & Marin, 1996).

Fish genomes typically featured dinucleotide AT/TA microsatellite repeats, a pattern observed in the genome of *T. tambroides* as well (Lei et al., 2021). However, the second most common dinucleotide repeat found in the genome was AC/GT, which aligns with the distribution found in *Danio rerio*, a species with more (AT) repeats than (AC) repeats (Chong et al., 2011). Nonetheless, other species like *Dichotomyctere nigoviridis*, *Oryzias latipes* and *Lates calcarifer*, exhibited a greater distribution of AC repeats than AT repeats, potentially due to their similar order as *D. rerio* and *T. tambroides* (Chong et al., 2011). Additionally, rare trinucleotide repeats (CGG)_n (16 counts) and (ACG)_n (2572 counts) were observed in *T. tambroides*, likely attributable to the presence of highly mutable CpG dinucleotides within the motif (Lei et al., 2021). The most prevalent trinucleotide repeats in *T. tambroides* are (AAT)_n repeats (17260 counts; 28.78%) due to their propensity to form hairpin structures, which can stabilize strand slippage (Chong et al., 2011; Moore et al., 1999). Furthermore, *T. tambroides* displays tetranucleotide repeats with lower G+C content, potentially impacting the mutation rate. Eckert et al. (2002) reported no statistically significant difference between repeats with 25% G+C content, but each of these differed significantly from repeats with 50% G+C content.

In this study, repetitive elements detected within the genome of *T. tambroides* accounted for 39.84% of the total genome content (Table 2.3). Previous studies have proposed a positive correlation between the expansion of repetitive elements and an increase in fish genome size, as indicated by research by Yuan et al. (2018). However, recent findings from a study examining SSRs across 14 fish species challenged this notion (Lei et al., 2021). It is suggested that a high content of repetitive elements can expedite the generation of novel

genes for adaptation, as observed in salmon, potentially linked to genome duplication events (Lien et al., 2016). Moreover, the variability found within microsatellites may stem from various selective constraints favoring certain types of microsatellites (Lei et al., 2021). Nonetheless, an overabundance of repetitive elements could lead to abnormal recombination and splicing, potentially causing genome instability (Yuan et al., 2018). In conclusion, the role of repetitive elements in genomic function remains somewhat ambiguous, and their impact on the genome may be shaped by specific environmental selection pressures, albeit with limitations up to certain levels (Yuan et al., 2018).

2.4.2 Accumulation of Repetitive Elements in Shaping *T. tambroides* Genome

Researchers have uncovered a potential link between the proliferation of repetitive elements and the observed augmentation in genome size among 52 teleost species (Yuan et al., 2018). Transposable elements (TEs) or transposons were previously regarded as molecular parasites as they are segments of genetic materials ensuring their own replication without the help of the host. Besides, due to their ability to leap around chromosomes in both prokaryotes and eukaryotes, they are sometimes known as ‘jumping genes’ as well (Kleckner, 1981). TEs were initially published in the 1950’s by Barbara McClintock (Pontarotti, 2017). Based on their ability to jump, TEs were classified into RNA transposons or retrotransposons and DNA transposons. The movement of RNA transposons is depending on RNA intermediate and reverse transcriptase to produce complementary DNA while DNA transposons depend on DNA intermediate and transposase to jump (Bourque et al., 2018). TEs can be categorized into five primary groups according to their enzymatic properties, sequences, and structure: long terminal repeat (LTR) retrotransposons, non-LTR retrotransposons (which encompass long and short interspersed nuclear elements: LINEs and

SINEs), cut-and-paste DNA transposons, rolling circle DNA transposons (known as Helitrons), and self-synthesizing transposons (referred to as Polintons) (Kapitonov & Jurka, 2008).

Initially, the discovery of TEs does not attract much scientific attention as it is wrongly treated as ‘junk DNA’. However, the important role of TEs in shaping host genomic landscape, rearranging chromosomes, speciation and sex determination as well as triggering individual genes evolution had been emphasized (Yuan et al., 2018; Shao et al., 2019; Gao et al., 2020). Therefore, assessing TEs is crucial for understanding the dynamics of genome evolution.

Among all the three mechanisms, cut-and-paste transposons is the most diverse containing at least 10 superfamilies identified in eukaryotes. They are distinguished by the presence of transposase encoded either autonomously or by inverted terminal repeats (ITRs) (Feschotte & Pritham, 2007). One of the most widely recognized superfamilies of cut-and-paste transposons is Tc1/*mariner*, known as the most prevalent DNA transposon in nature. It earned its name from the founding gene discovered in *Caenorhabditis elegans* (Transposon *C. elegans* number 1) and features transposase flanked by ITRs (Plasterk et al., 1999). Studies focusing on sequence analysis and phylogenetic distribution have suggested potential horizontal transmission of Tc1/*mariner* between fish and frogs, as well as between fish and lampreys (Kuraku et al., 2012).

Although an abundance of repetitive elements within the genome may expedite the creation of new genes for adaptation purposes, an excessive presence can lead to irregular recombination and splicing, potentially causing genome instability (Hong, 1998). Thus, it is not advisable for repetitive elements to proliferate unchecked within the genome, but rather

to be constrained within certain limits and influenced by specific natural selection pressures (Yuan et al., 2018).

In this study, TEs comprised 39.84% of the repetitive elements in *T. tambroides*, a proportion likely linked to its genome duplication event (Yuan et al., 2018). This discovery aligns with the prevalence of repetitive elements observed across species within the order Cypriniformes, such as *Danio rerio*, *Cyprinus carpio*, *Sinocycloheilus graham*, *Sinocycloheilus rhinoceros*, *Sinocycloheilus anshuiensis*, and *Pimephales promelas*, where the documented range of repetitive elements falls between 35% and 40% (Yuan et al., 2018). The prevalence of class II transposons (20.33%) in *T. tambroides* genome exceeded that of class I transposons (4.31%). This pattern may be attributed to the higher occurrence of class I transposons in cartilaginous fish and lampreys compared to bony fishes (Yuan et al., 2018). However, the reason for these differences is still lacking scientific explanation. One potential explanation for the class I transposons accumulation within ancient cartilaginous fish and lamprey could be linked to their evolutionary adaptations, as these transposons are implicated in the regulation of gene expression (Brosius, 1999; Gess et al., 2006). Furthermore, Yuan et al. (2018) noted the lesser appearance of class I transposons within the genome of bony fish genomes than those in cartilaginous fish and lamprey, possibly due to the putative mechanism to counteract the invasiveness of RNA on respective genomes. Moreover, the expansion of TC1-IS630-Pogo elements, which fall under the DNA transposons category is believed to be the main driver of TE expansion in *Corydoradinae* as its abundance increases (Marburger et al., 2018).

Previous research has highlighted the relationship between categories and the proportion of repetitive elements of fish and their living environment (Yuan et al., 2018). The findings of Yuan et al. (2018) support the notion that freshwater fish exhibit a higher

prevalence of class II transposons compared to marine fish. This enrichment of repetitive elements, which correlates with the host's living environment, suggests an integral role of these elements in genomic evolution and potential facilitation of adaptation to respective habitats (Yuan et al., 2018). Moreover, environmental stressors like floods and droughts in freshwater ecosystems can elevate transposition rates, leading to the generation of new genetic variants that enhance host adaptation to the environment (Schrader et al., 2014). Consequently, this phenomenon may elucidate why freshwater ray-finned fish possess a smaller effective population size and larger genome size compared to marine species, contributing to the complexity of their genomes (Schrader et al., 2014). Several studies have emphasised the need to explore on the expansion of class II transposons, but this can be challenging due to the possibility of vertical and horizontal transfer (Abrusan et al., 2006).

2.4.3 Immunity-related Annotation

The TLR family represents a crucial category of pattern-recognition receptors expressed on antigen-presenting cells within the fish immune system. These receptors play a pivotal role in innate immune responses and contribute to the promotion of adaptive immunity (Akira et al., 2006). Within the genome of *T. tambroides*, a total of 11 interleukins (IFNs) and 14 *TLR* genes were successfully identified (Table 2.6). To date, research has identified at least 21 types of TLRs in fishes, including TLR1-5, 5S, TLR7-9, TLR13, TLR14, TLR18-23, and TLR25-28. These encompass both orthologs of mammalian TLRs and species-specific TLRs (Nie et al., 2018). Vertebrate TLRs are commonly categorized into six main subfamilies: TLR1, TLR3, TLR4, TLR5, TLR7, and TLR11 (Roach et al., 2005). In the genome of *T. tambroides*, both TLR1 and TLR2 have been identified within the TLR1 subfamily. TLR1 functions to recognize triacylated lipoproteins as well as

mycobacterial products by forming a heterodimer with TLR2 in mammals (Randelli et al., 2008). These TLRs have previously been characterized in other fish species including zebrafish (*Danio rerio*), common carp (*Cyprinus carpio*), and grass carp (*Ctenopharyngodon idella*) (Jault et al., 2004; Fink et al., 2016; He et al., 2016). The TLR2 signaling pathway has been linked to the recognition of probiotics, particularly *Psychrobacter* sp., and the subsequent activation of the mucosal immune system in orange-spotted grouper (Nie et al., 2018). Conversely, TLR14 is known to play a role in immune responses against various pathogens, including both Gram-negative and Gram-positive bacterial infections, as well as viral infections (Hwang et al., 2011).

TLR3 plays a crucial role in detecting various molecules such as double-stranded RNA (dsRNA), endogenous cellular mRNA, and sequence-independent small interfering RNAs, thereby contributing to innate immune responses (Sahoo et al., 2012). When dsRNA binds to TLR3, it triggers the production of interferons and inflammatory cytokines in fish cells. On the other hand, members of the TLR4 family are primarily involved in recognizing lipopolysaccharide (LPS) and are renowned as the most well-known pattern recognition receptors (PRRs) (Lu et al., 2008). During grass carp reovirus (GCRV) infection in grass carp, TLR4 upregulation has been observed in both the muscles and liver, indicating its involvement in immune function (Huang et al., 2012). Within the genome of *T. tambroides*, a member of the *Cyprinidae* family, TLR4 has been detected. This discovery partially supports the emergence of the TLR4-lipopolysaccharide (LPS) signaling pathway after the evolutionary split between fish and tetrapods (Nie et al., 2018), contrasting with the absence of TLR4 in pufferfishes (*Takifugu rubripes* & *Takifugu niroviridis*) and stickleback (*Gasterosteus aculeatus*) (Oshiumi et al., 2003; Li et al., 2012). TLR5 as the only member of TLR5 subfamily, was found in both membrane-bound (TLR5M) and soluble non-

transmembrane forms (TLR5S) (Tsujita et al., 2004). TLR5 recognizes both pathogen-associated molecular patterns and damage-associated molecular patterns, leading to activation of the MyD88-dependent TLR signalling pathway during instances of heat and cold shock. Basu et al. (2015) observed such a phenomenon within Indian major carp. In teleost, TLR5M exhibits significant expression across tissues like the head kidney, spleen, liver, and brain, whereas TLR5S predominantly appears in the liver (Bai et al., 2017).

The TLR7 subfamily, which encompasses TLR7, TLR8, and TLR9, shares structural similarities with mammalian homologs. Meijer et al. (2004) observed reduced TLR7 expression in zebrafish blood and lymphoid tissues but noted an increase in TLR8 expression following an 8-week infection with the pathogen *Mycobacterium marinum*. Moreover, after *Vibrio parahaemolyticus* infection, an increased TLR9 expression was detected in spleen and kidney tissues in common carp (Kongchum et al., 2011), large yellow croaker (Yao et al., 2008), and gilthead sea bream (Franch et al., 2006). The TLR11 subfamily comprises several fish-specific TLRs. For instance, the challenge with *Aeromonas hydrophila* resulted in upregulated TLR22 expression in various tissues of common carp, suggesting its significant role in systemic and mucosal defense against viral or bacterial infections (Wang et al., 2017).

The major histocompatibility complex (MHC) has a vital role in identifying foreign substances. The complex functions by displaying pathogen peptide fragments for the subsequent removal by T cells (Neefjes et al., 2011). The genome data includes both MHC class I and II. The high variability of MHC genes makes them promising markers for disease resistance in teleost fish (Lanfords et al., 2001). Both TLR and MHC genes were found present within the transcriptome of *Trachinotus ovatus* (Zhenzhen et al., 2014) and further investigations into these genes could aid the scientific community in having a better

understanding of the immune system of *T. tambroides* as well as other teleost. However, a previous study noted a lack of functional specific antibody responses in MHC complex II and CD4⁺ T cells in the Atlantic cod genome, potentially attributed to adaptive strategies developed to counter infectious diseases, such as the significant expansion of specific genes like Toll-like receptors, antimicrobial peptides, and MHC I (Malmstrøm et al., 2016).

2.4.4 Growth-related Annotation

According to Zhenzhen et al. (2014), the growth process is intricate, encompassing the regulation of appetite, protein and lipid metabolism, weight gain, and muscle development. Within the genome of *T. tambroides*, several genes associated with appetite and muscle growth were identified, including neuropeptide Y (*NPY*), pro-opiomelanocortin (*POMC*), leptin, and myostatin (Table 2.7). These genes serve diverse functions, with *NPY* and *POMC* exhibiting antagonistic roles in stimulating and inhibiting feeding, respectively (Zhenzhen et al., 2014). Additionally, leptin is known to regulate energy intake and expenditure (Zhenzhen et al., 2014), while myostatin acts as a negative regulator of muscle growth, with its polymorphism linked to growth traits (Nadjar-Boger & Funkenstein, 2011).

The genome of *T. tambroides* contains genes related to both protein and lipid metabolism. Among these genes are chymotrypsin-like elastase (*cela*), cathepsin (*clsL*), and cathepsin K (*clsK*), which are involved in protein metabolism. Additionally, lipid metabolism-related genes, including lipase C (*lipC*), phospholipase A2 (*pla2*), elongation of very long-chain fatty acid family 6 (*elovl6*), apolipoprotein B (*apob*), acetyl-CoA carboxylase (*ACACA*, *ACACB*), and fatty acid synthase (*FASN*), were also identified. Given that *T. tambroides* is classified as a semi-fatty fish with a considerable amount of muscle crude lipid (4.6-5.2%), lipid metabolism plays a vital role in this species. Lipase and

apolipoprotein are types of lipid-associated proteins crucial for lipid hydrolysis and maintaining lipid homeostasis by facilitating the transport of triacylglycerol and phospholipid from the liver to other tissues.

ACACB plays a dual role in determining fat yield and percentage, as well as contributing to protein yield (Han et al., 2018). Meanwhile, the expression of *FASN*, a crucial component, varies depending on the fatty acid content found in both fat and meat (Renaville et al., 2018). Furthermore, the presence of glucose-6-phosphatase (*g6pc*), involved in regulating carbohydrate metabolism, was also noted in *T. tambroides*. These genes, involved in growth, appetite, protein and lipid metabolism, and carbohydrate metabolism, could potentially serve as molecular markers for growth-related traits in *T. tambroides* breeding programs. Nonetheless, further investigation is needed to elucidate the specific roles of these genes in the growth of *T. tambroides*.

CHAPTER 3

ISOLATION, IDENTIFICATION AND CHARACTERISATION OF *P. KOREENSIS* ISOLATED FROM DISEASED EMPURAU (*T. TAMBROIDES*)

3.1 Introduction

As highlighted by the FAO (2016) and the DOF (2019), aquaculture holds significant importance as a primary source of animal protein in diets worldwide. With the world population rapidly expanding, by 2030, there is a pressing need to increase aquaculture production by an extra 46.4 million metric tons to ensure sustainable food production and security (World Bank, 2013). Malaysia stands out with a fish product demand of 59 kg per capita in 2016, one of the highest in the world according to the FAO (2020a). Consequently, Malaysia heavily relies on its seafood production sector to meet its protein needs, with the aquaculture sector alone contributing up to 20% of total seafood production (FAO, 2020b). Since 1980, the industry has experienced steady growth, culminating in an output of 1.69 million tons in 2017, as documented by Waiho et al. (2020).

Empurau (*Tor tambroides*) (Bleeker, 1854) holds significant value among riverine fishes in Southeast Asia, renowned not only for its culinary appeal but also for its ornamental and sport fishing significance. According to Ingram et al. (2005), with its exceptional nutritional content, distinctive flesh texture, and flavor, it commands top prices in fish markets, fetching up to RM 400 per kg. Belonging to the *Cyprinidae* family inhabiting freshwater bodies, *T. tambroides* shares a biogeographical range with *T. douronensis* as well as *T. tambra* in Indonesia and Malaysia (Pinder et al., 2019). However, the burgeoning industry faces substantial challenges, notably disease outbreaks that demand immediate

attention. Fish farms are plagued by persistent viral and bacterial infections (Chiew et al., 2019), posing significant economic risks to the sector.

Pseudomonas koreensis has been isolated from various sources, including Korean agricultural soil (first isolation) (Kwon et al., 2003), rhizophores (Lozano et al., 2017), agricultural soils (Lopes et al., 2018), gold mining soil (Babalola & Ayangbenro, 2019), yak milk (Kaur et al., 2019), horse chestnut trees (Crovadore et al., 2020), and bronchiectasis patients (Fluit et al., 2020). Additionally, it has been identified as a fish pathogen in *Tor putitora* (golden mahseer) (Shahi & Malik, 2014). While the environmental *P. koreensis* strains discovered in soil are extensively studied for their role in promoting plant growth, fish pathogenic strains have received limited examination. However, there is increasing concern about *Pseudomonas* infections in public health, particularly through the consumption of raw fish and its byproducts, leading to serious food-borne illnesses (Zilberberg & Shorr, 2012). Previous reports have linked fish-isolated *Pseudomonas aeruginosa* to hospital-acquired pneumonia in humans (Novotny et al., 2004). Furthermore, cases have been documented of *P. koreensis* infections in patients with bronchiectasis (Fluit et al., 2020), and also causing mixed infectious keratitis alongside *Aspergillus fumigatus* (Khoo et al., 2021), indicating its potential threat to infect human hosts.

Clinical manifestations reported in infected fish include eye opaqueness, hemorrhagic septicemia, necrosis of the gills, distension of the abdominal, congestion in the kidneys, and liver friability (Ardura et al., 2014; Shahi & Malik, 2014). However, the manifestation of these symptoms may vary and is not consistently observed in all infected fish; the pathogenicity differs depending on the specific bacterial isolates or strains (Chen et al., 2019a). Hence, it is essential to explore characteristics including virulence, growth patterns and histological lesions associated with the pathogen. Global reports have

highlighted the extensive use of antimicrobial agents in the aquaculture sector and the transmission of multidrug-resistant bacterial strains between terrestrial and aquatic ecosystems (Cabello, 2006).

Assessing virulence factors is crucial for evaluating the virulence and toxicity of bacterial pathogens (Chen et al., 2019a). These virulence genes, including those related to adherence (*fliC*, *pilB*), antiphagocytosis (*algD*), quorum sensing (*hdtS*), protease (*aprA*), regulation (*gacA*, *gacS*), secretion system (*phnB*), iron uptake (*pvdA*, *pchE*), and toxin production (*exlA*, *hcnA*), play key roles in the pathogenesis and epidemiology of fish-isolated *P. koreensis*. However, limited data are available regarding these factors, highlighting the importance of their survey in understanding this species. Additionally, studying the growth characteristics of the bacterial isolate under various conditions is essential for identifying suitable growth environments for this pathogen. The main objective of the current study is to characterize the virulence, antibiotics susceptibility and growth characteristics of isolated *P. koreensis* from the local aquaculture farm.

In this chapter, the dominant bacterium isolated from diseased *Tor tambroides* obtained from a local aquaculture farm is identified as *P. koreensis* CM-01 strain. The chapter presents the results of the isolation, identification, pathogenicity, virulence genes, and growth characteristics of *P. koreensis*.

3.2 Materials and Methods

3.2.1 Bacterial Isolation

P. koreensis was isolated from the diseased adult *T. tambroides* (intestine and liver) sampled from a local aquaculture farm (GPS coordinates: 1° 32' 53.647" N, 110° 32' 53.233" E). Aeromonas isolation agar (Sigma Aldrich; US) was used for the isolation of *P. koreensis* with the addition of Ampicillin as a selective supplement (0.5mg/mL). Cultured plates were incubated at 30 °C for 24 hr. The colonies that appeared on the agar plates were sub-cultured and purified until dominant colonies were observed. A single colony was selected and inoculated in LB broth at 30 °C for 18hr. Then, the main culture was preserved in a final concentration of 20% sterile glycerol (v/v) at -20 °C (short-term storage) and -80 °C (long-term storage). A single colony was picked and added as the template for colony PCR for its identity confirmation through 16S rRNA gene sequencing. A pair of universal 16S primers (27F & 1492R) was used. After 16S rRNA sequencing, the sequences were identified through a BLAST search via NCBI.

3.2.2 Physiological and Biochemical Characterisation

The isolated bacteria, *P. koreensis* was characterized through its physiological and biochemical properties. The strain was grown on different agar, includes LB agar (BD Difco™; US), Aeromonas isolation agar (Sigma-Aldrich; US), Nutrient agar (Hi-Media; India), EMB agar (Hi-Media; India) and MacConkey agar (Hi Media; India), to observe its growth and colony appearance (Figure 3.1). Gram staining was carried out to observed on the morphologies of the isolated bacteria and to confirm on the Gram-negative or Gram-positive property of the bacteria (Figure 3.2). API 20E system was used to measure the physiological and biochemical characteristics of *P. koreensis*, including maltose, glucose, sucrose, lactose,

fructose, raffinose, xylose, melezitose, sorbose, L-rhamnose, sorbitol, inositol, dulcitol, adonitol, urea, esculin, Voges-Proskauer test, tartrate, malonate, citrate, starch, salicin, H₂S, indole, lysine decarboxylase, ornithine decarboxylase, arginine dihydrolase and motility.

3.2.3 16S rRNA Gene Sequence Analysis

Total genomic DNA of the isolated strain CM-01 was extracted via modified CTAB-based protocol (Chung, 2018). The 16S rRNA was amplified via PCR (Biorad; California) using 16S universal primer: 27F: 5' 3' and 1492R: 5' 3'. Amplification conditions were set as follow: initial denaturation 95 °C for 3 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s and extension at 72 °C for 1 min and subsequently final extension at 72 °C for 5 min. The amplified PCR products were visualized on 1.5% agarose gels by staining with ethidium bromide. The PCR products were purified with a GEL/PCR Purification Mini Kit (Favorgen Biotech Co.) (Figure 3.3) and then sent for sequencing at Apical Scientific Sdn Bhd (Selangor, Malaysia). A BLAST search for sequences was performed through NCBI website. A phylogenetic tree was generated to infer its phylogenetic relationship with the *Pseudomonas* species (Figure 3.4).

3.2.4 Growing Characteristics

Next, the growing characteristics of the isolated bacteria were determined. The impact of pH (ranges from 3 to 11) and temperature (assessed at 27 °C, 30 °C and 42 °C) on the growth of isolated *P. koreensis* were identified (Figure 3.9) To achieve this, isolated *P. koreensis* were cultured in LB broth until its OD 600 of the suspension reaches 1.2. All the

bacterial suspension were cultured at 150 rpm and the growth were observed at 4-hour intervals at OD 600 for 52 hr.

3.2.5 Drug Sensitivity Tests

CM-01 strain was sub-cultured and tested for an expanded range of antimicrobial sensitivities using the VITEK2 system (bioMérieux; France). Specifically, the VITEK2 AST-N314 cards were used to test and the antibiotics included were cefotazime (1-64 µg/mL), ceftazidime (1-32 µg/mL), cefepime (1-64 µg/mL), ciprofloxacin (0.25-4 µg/mL), imipenem (0.25-16 µg/mL), meropenem (0.25-16 µg/mL) and piperacillin/tazobactam (4/4-128/4 µg/mL). The susceptibility category was designated according to the 2023 Clinical and Laboratory Standards Institute (CLSI) M100-Ed33 guidelines. *Pseudomonas aeruginosa* ATCC 27853 was used as the quality control strain for the antibiotics susceptibility testing (Table 3.4).

3.2.6 Screening of Virulence Genes

Next, isolated bacteria were screened for the presence of virulence genes. The virulence genes were selected from the list of 150 virulence genes predicted from the whole genome data of *P. koreensis* using Virulence Factors Database (<http://www.mgc.ac.cn/VFs/>) reported previously (Kho et al., 2023). Twelve virulence genes were screened, including *fliC*, *algD*, *pilB*, *phnB*, *pchE*, *hcnA*, *hdtS*, *exlA*, *aprA*, *gacS*, *pvdA* and *gacA* through conventional PCR (Figure 3.8). Each PCR reaction was commenced with denaturation at 94 °C for 2 min, then 30 cycles of amplification, and finally extension at 72 °C for 10 min. Each cycle included 94°C denaturation for 30 s, annealing for 50 s and 72 °C extension for 30 s.

Table 3.1 summarised on the PCR reaction components used for the virulence gene amplification while Table 3.2 showing the amplicon size, primer sequences and annealing temperature of each target gene involved in this study.

Table 3.1: PCR reaction components for virulence genes amplification from *P. koreensis*

Component	Final Concentration	Final Volume (µL)
5X Green GoTaq Reaction buffer	1X	5
10 mM dNTP	0.2 mM	0.4
25mM MgCl ₂	1.0 mM	0.8
10 µM Forward primer	0.5 µM	1.0
10 µM Reverse primer	0.5 µM	1.0
GoTaq DNA Polymerase (5u/µL)	1 unit	0.25
DNA template	-	1
Ultrapure water		10.55
Final volume		25.0

Table 3.2: PCR targeted virulence genes, targeted amplicon size and annealing temperatures of CM-01 strain

Target Gene	Product size (bp)	PCR primers sequence (5' – 3')	Annealing temperature (°C)	References
<i>fliC</i>	680	F: TTCAGTACAGCGGAAGGCAG R: ATGCAATCGCAAATCCGTGG	62	Feng et al., 2015
<i>algD</i>	399	F: TTCGAGAATCGGGATCACCAC R: GGTTTGGGTTACGTTGGCG	58	Benie et al., 2017
<i>pilB</i>	702	F: GTCGATTTCCAGCGAGTTGC R: ATTGATGCGCTGGGTTACGA	62	Benie et al., 2017
<i>aprA</i>	669	F: ACCTCACGGTCAATGGCAAA R: CGTAGAGCTTCTGGATCGCC	56	Bardoel et al., 2011
<i>gacA</i>	434	F: TATCGATGGCTTGCAGGTGG R: TCTGCACCTTCTGACAACCG	58	O' Malley et al., 2020
<i>gacS</i>	743	F: GTCCGTTCCGGTCCTTGATT R: CTCGATGTCACGCCCTTCAA	60	O' Malley et al., 2020
<i>pchE</i>	678	F: TCTGGCTGAAGGCGTATTCG R: TGACCTACGCCGAACCTCAAC	62	Andreozzi & Uhlich, 2020
<i>phnB</i>	432	F: GTTCAAACATCTCCATCGCCC R: CCGAGGGGTATCACAGCATT	62	Mahajan-Miklas et al., 1999
<i>hcnA</i>	291	F: GCACTGCCTAGAAAGAACCTTC G R: ATGCGGTTGACCCGGGTTTC	62	Pessi & Haas, 2000
<i>hdtS</i>	346	F: CGGCATTGTGTGCAATCGG R: CCCTGCGTGATTCAGTCGAA	62	Hosoe et al., 2022
<i>exlA</i>	723	F: CAACGGCATTTCGGTGAACG R: GTTTCACCTGTCGGTCTGGGT	62	Reboud et al., 2017
<i>pvdA</i>	667	F: TGGATGTGCTGTTCTCGAC R: AACTCCGGCGAGAACACTTC	62	Fazeli & Momtaz, 2013

3.2.7 Experimental Infections

T. tambroides provided by LTT aquaculture Sdn Bhd were raised at the Faculty of Resource Science and Technology, Universiti Malaysia Sarawak. The fish were maintained at optimal rearing conditions for one week prior to experimental infection. Aeration was provided for DO maintenance and fed with commercial feeds once daily. The fish were reared in treated tap water and twice water changes (40% of the total water volume) were done per week.

After the identity of the isolated bacteria had been confirmed, the bacteria were prepared in five bacterial concentrations: 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 and 1×10^9 CFU/mL. A total of 60 healthy *Tor tambroides* were randomly divided into 6 groups with each tank containing 10 fish to test on the aforementioned bacterial concentration along with a control treatment ($1 \times$ PBS). The dose of both infection and control was 200 μ L/fish. After introduction of *P. koreensis* into the fishes via injection, all fishes were raised for 7 days at 25 °C for observation and record of pathological symptoms. Pathogenicity was confirmed through the occurrence of the disease, include clinical signs (Figure 3.6). The sacrificed fish was dissected and bacteria from the liver and spleen were re-isolated. The identity of the isolated bacteria was confirmed through 16S rRNA amplification. Mortalities were summarized in Table 3.3 while its LD₅₀ dosage being calculated using Miller and Tainter method (1980) (Figure 3.5).

3.2.8 Histopathological Observations

Subsequently, for histopathological observation, the intestine, gills, liver, spleen, kidney and stomach were dissected from the experimentally infected fish (10^7 and 10^8 cfu/ml infection group). Next, the isolated tissue were fixed with 4% paraformaldehyde (PFA) at 4

°C for 48 hours. The fixed tissues were washed in PBS at least 3 times for 20 mins to remove all the possible carried-over PFA. Next, the tissues were placed in 15% sucrose in PBS (4 °C) followed by 30% sucrose in PBS (4 °C) until the tissues sunk in both solutions for cryopreservation and to prevent cell deformation. The tissues were then embedded into an agar-sucrose solution (1.5% agar with 5% sucrose) and left aside for 5 to 10 mins for agar solidification. The agar blocks were trimmed at least two to three mm around the tissue. Next, place the trimmed agar block into 30% sucrose solution at 4 °C until the agar block has sunk. Subsequently, tissue freezing medium was used as a slightly raised platform on the cryostat chuck to position the agar block onto the chuck. The tissues were sectioned at 5 µm at -35 °C and desired sections were picked up with a room temperature glass slide. The sections were eventually stained with hematoxylin and eosin (H&E) for histopathological observation. Prior to H&E staining, the sections were fixed rapidly with methanol for 1 min and rinsed with water. Next, the sections were stained with hematoxylin regressively for 10 min and rinsed with tap water again. The differentiation step was performed by adding acid alcohol onto the sections and followed by rinsing in tap water. Prior to eosin staining, the sections were rinsed with 90% ethanol 3 times as eosin is mostly made up of ethanol. Next, the sections were stained with eosin for 2 min. The excess eosin was tipped off and followed by dehydration step through the addition of 90% ethanol for 3 times, 95% ethanol and absolute ethanol. The sections were added with xylene for 2 times with 10 mins each. Finally, two drops of 50% glycerol solution were added to the sections for mounting. After the coverslip was slowly mounted onto the sample, the sections were observed under a light microscope under 40X and 100X magnification. The same organs were collected from healthy fish as an observational histopathological comparison. The results were shown in Figure 3.7.

3.3 Results

3.3.1 Morphological and Biochemical Characteristics

The growth of *Pseudomonas koreensis* CM-01 strain was observed via growth on various agar mediums, such as Aeromonas isolation agar, Luria Bertani (LB) agar, Eosin-Methylene blue (EMB) agar, MacConkey agar, and Nutrient agar. Next, the cultured bacteria was incubated at 30 °C for 24 hours. The resulting colony growth patterns are depicted in Figure 3.1. On Aeromonas isolation agar, CM-01 exhibited green round colonies (Figure 3.1A), while white round opaque colonies were observed on LB agar (Figure 3.1B). Colorless colonies were noted on EMB agar (Figure 3.1C) and MacConkey agar (Figure 3.1D), indicating that CM-01 is a non-lactose fermenter. Additionally, white and opaque colonies were noted on nutrient agar (Figure 3.1E). Gram staining revealed pink-colored rod-shaped bacteria arranged singly (Figure 3.2). Further analysis of biochemical properties using the API 20E system revealed positive results for citrate utilization, arginine dihydrolase, gelatinase, acetyl methyl carbinol, and oxidase activity. However, CM-01 tested negative for the remaining 16 tests, including β -galactosidase, lysine decarboxylase, ornithine decarboxylase, H₂S production, urease, tryptophan deaminase, indole production, and the assimilation of various sugars such as glucose, mannose, inositol, sorbitol, rhamnose, sucrose, melibiose, amygdalin, and arabinose. CM-01 was revealed as a species belonging to genus *Pseudomonas* based on the 7-digit profile number (2203004) via the apiweb™ identification software ([https:// apiweb.biomerieux.com](https://apiweb.biomerieux.com)).

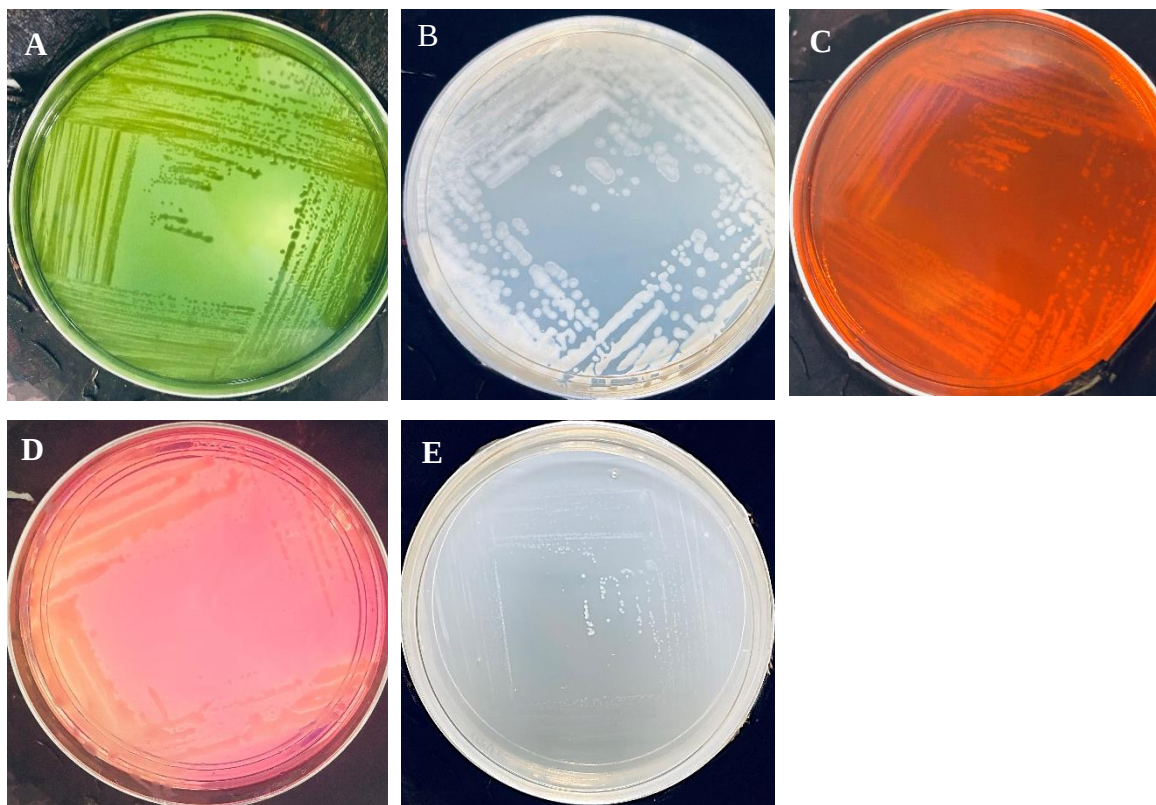


Figure 3.1: Growth of CM-01 strain on: (A) Aeromonas isolation agar, (B) LB agar, (C) EMB agar, (D) MacConkey agar and (E) Nutrient agar after culturing at 30 °C for 24 hours

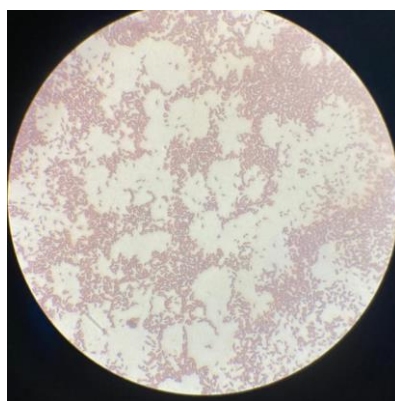


Figure 3.2: The Gram staining properties of the CM-01 strain with pink appearance observed

3.3.2 Sequence Analysis of the 16S rRNA gene

The 16S rRNA gene sequence obtained was approximately 1500 bp in length (Figure 3.3) (OR816052.1). A phylogenetic tree was constructed (Figure 3.4) via the amplified 16S rRNA sequence to elucidate the phylogenetic connection between the CM-01 strain and selected *Pseudomonas* species. In this tree, CM-01 clustered together with *Pseudomonas koreensis* BS3658, forming a distinct clade. After the experimental infection, for the re-isolation of *P. koreensis*, organs were collected from the fish sample and Aeromonas isolation agar was utilized for the cultivation of the bacteria. Subsequently, PCR was performed prior to sequencing to amplify the 16S rRNA gene from the isolated bacteria.

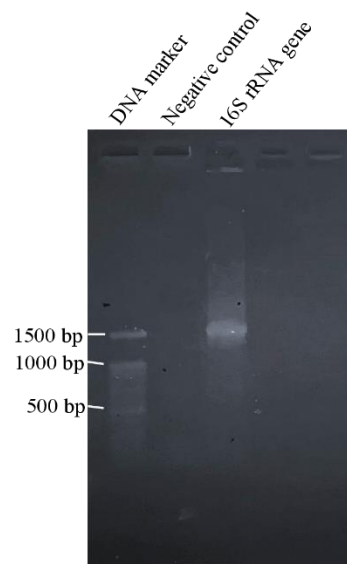


Figure 3.3: Agarose gel electrophoresis of PCR products of 16S rRNA gene of CM-01 with the size estimated through 100 bp DNA ladder (Transgene, French)

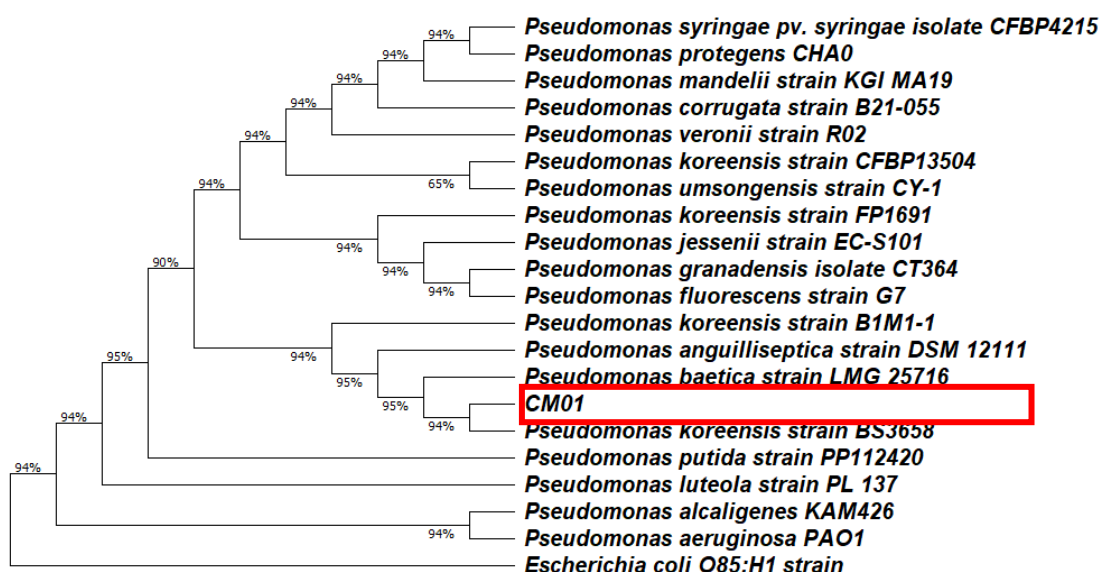


Figure 3.4: Neighbour-joining phylogenetic trees based on 16S rRNA gene of *Pseudomonas* strains using 1,000 bootstraps with a red box indicating the location of *Pseudomonas koreensis* CM-01 strain

3.3.3 Experimental Infections

Table 3.3 summarizes on the mortality rates of juvenile *T. tambroides* infected with the CM-01 isolate ($p < 0.01$). Significant differences were exhibited from almost all of the treatment groups as compared to the control group, excluding the 10^5 treatment group. Miller and Tainter method was used for the calculation of LD_{50} on Figure 3.5, which shown as 5.01×10^7 cfu/mL (Miller & Tainter, 1944).

The symptoms manifested in fish that were experimentally infected, are portrayed in Figure 3.6, which closely resemble those reported in natural illnesses back in 2022 (Lau et al., 2022). Some prominent clinical symptoms included internal fluid accumulation, loss of scales, coelomic distension, petechial hemorrhage, pop-eyes, and inflamed vent.

Table 3.3: Mortality of fish infected with CM-01 isolate across 7 days of experimental observation with the p-value shown from Tukey test in comparison with the control group

Concentration (cfu/ml)	Dose/ μ L	Number of tested fishes	No. of dead fish at different time			Mortality (%) (p-value)
			24 h	48 h	3-7 days	
10^9	200	10	7	10	10	100 (< 0.01)
10^8	200	10	5	6	6	60 ± 5.8 (< 0.01)
10^7	200	10	0	0	6	60 ± 5.8 (< 0.01)
10^6	200	10	0	0	1	10 ± 5.8 (< 0.05)
10^5	200	10	0	0	0	0 (no significant difference)
Control	200	10	0	0	0	0

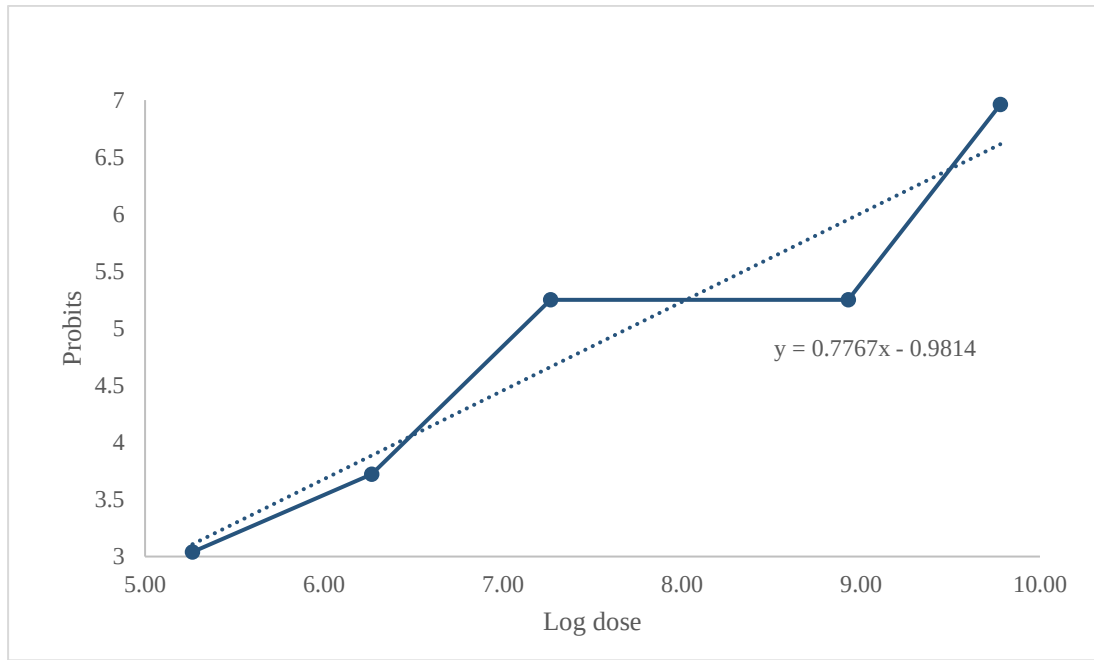


Figure 3.5: Log dose versus probit for the calculation of median lethal dosage of *P. koreensis* CM-01 in juvenile *T. tambroides*

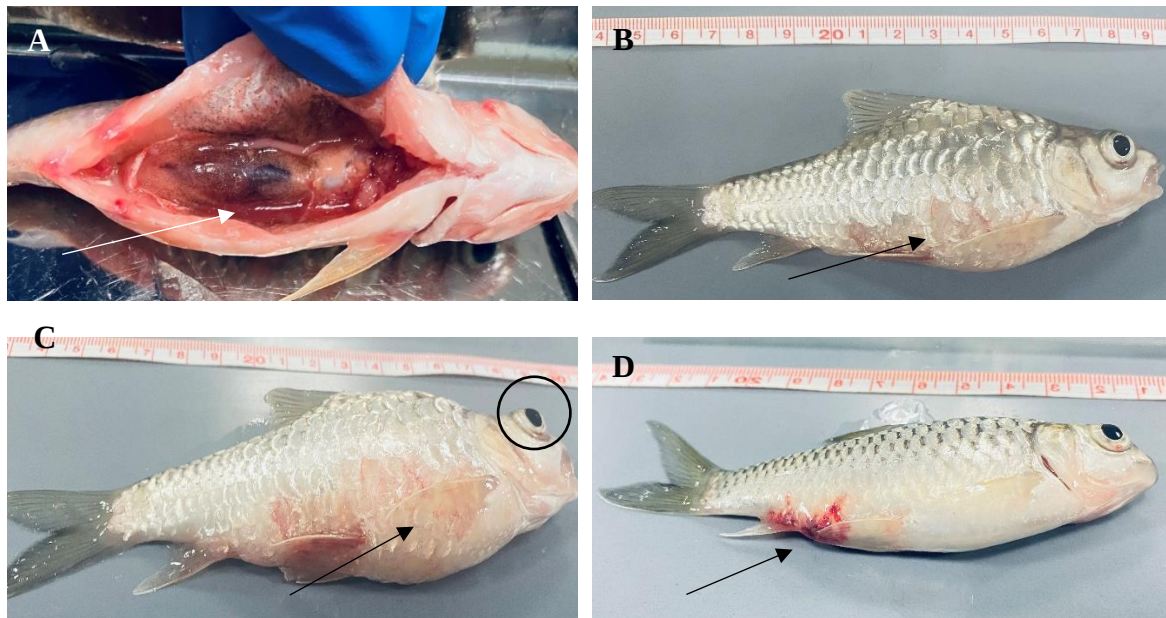
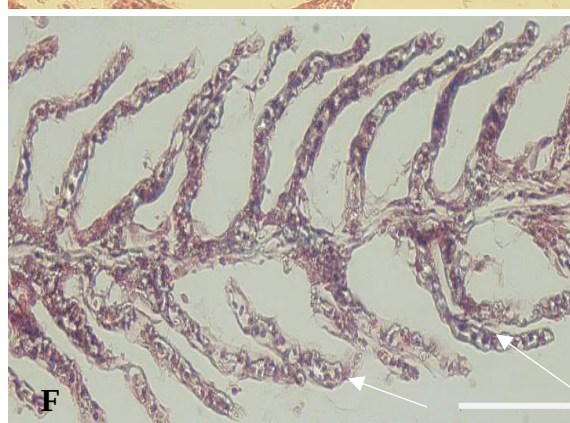
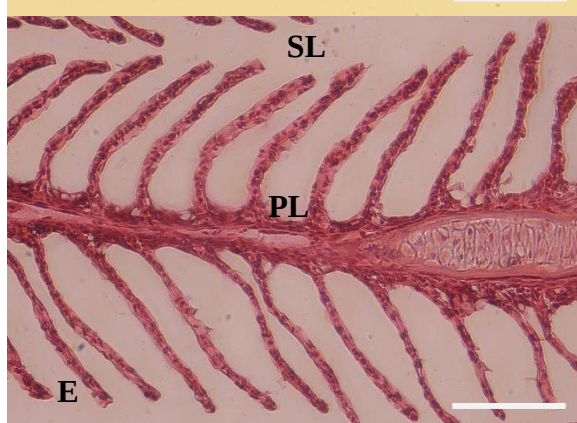
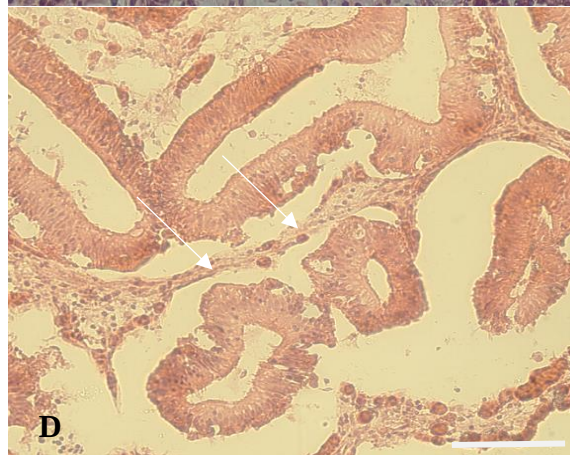
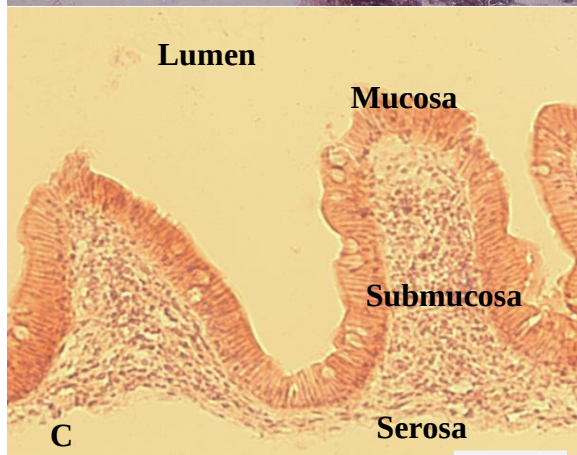
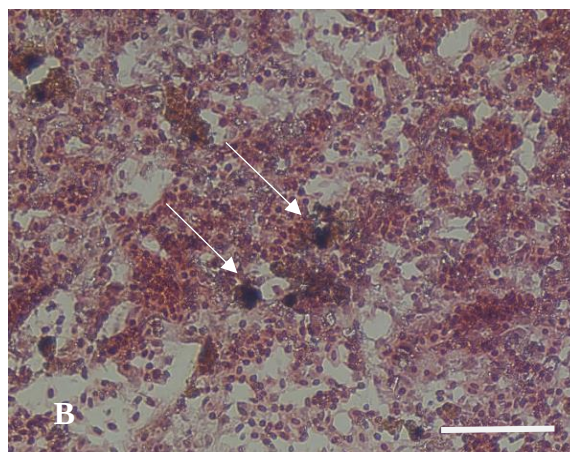
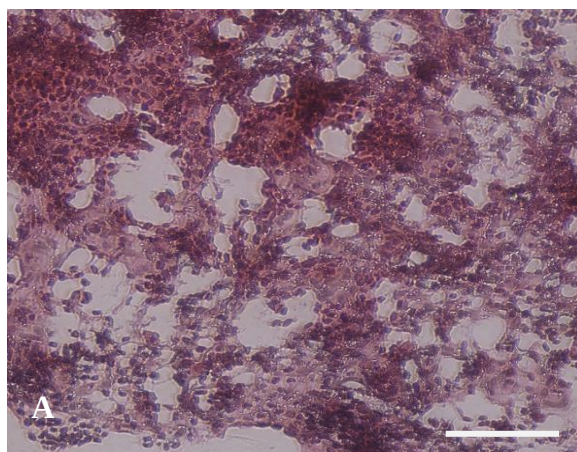


Figure 3.6: Pathological features of *Tor tambroides* infected with CM-01 strain. A: Internal fluid accumulation (white arrow); B: loss of scales (black arrow); C: coelomic distension (black arrow), petechial haemorrhage, pop eyes (circle); D: and an inflamed vent (black arrow)

3.3.4 Histopathological Observations

Histopathological observations had shown severe hemorrhages across a number of tissues within the diseased fish. The diseased fish was infected with LD₅₀ dosage of the strain while healthy fish acted as a control group to detect any histopathological differences following the experimental infection. The spleen of the diseased fish had shown severe splenitis accompanied by fibrinoid necrosis of the ellipsoidal sheath (Figure 3.7 A & B). Furthermore, on Figure 3.7 C & D, there were necrosis found on the epithelial layer of the intestinal lumen alongside with sloughed necrosis debris. Additionally, there were clubbing and thickening found across secondary lamellae (Figure 3.7 E & F) as well as interstitial congestion and renal tubular necrosis (Figure 3.7 G & H) in the disease fish. Besides, thin arrows on Figure 3.7 I & J show lymphocytes migrating from the dilated blood vessels and aggregating on the focal areas of the livers in the diseased fish.



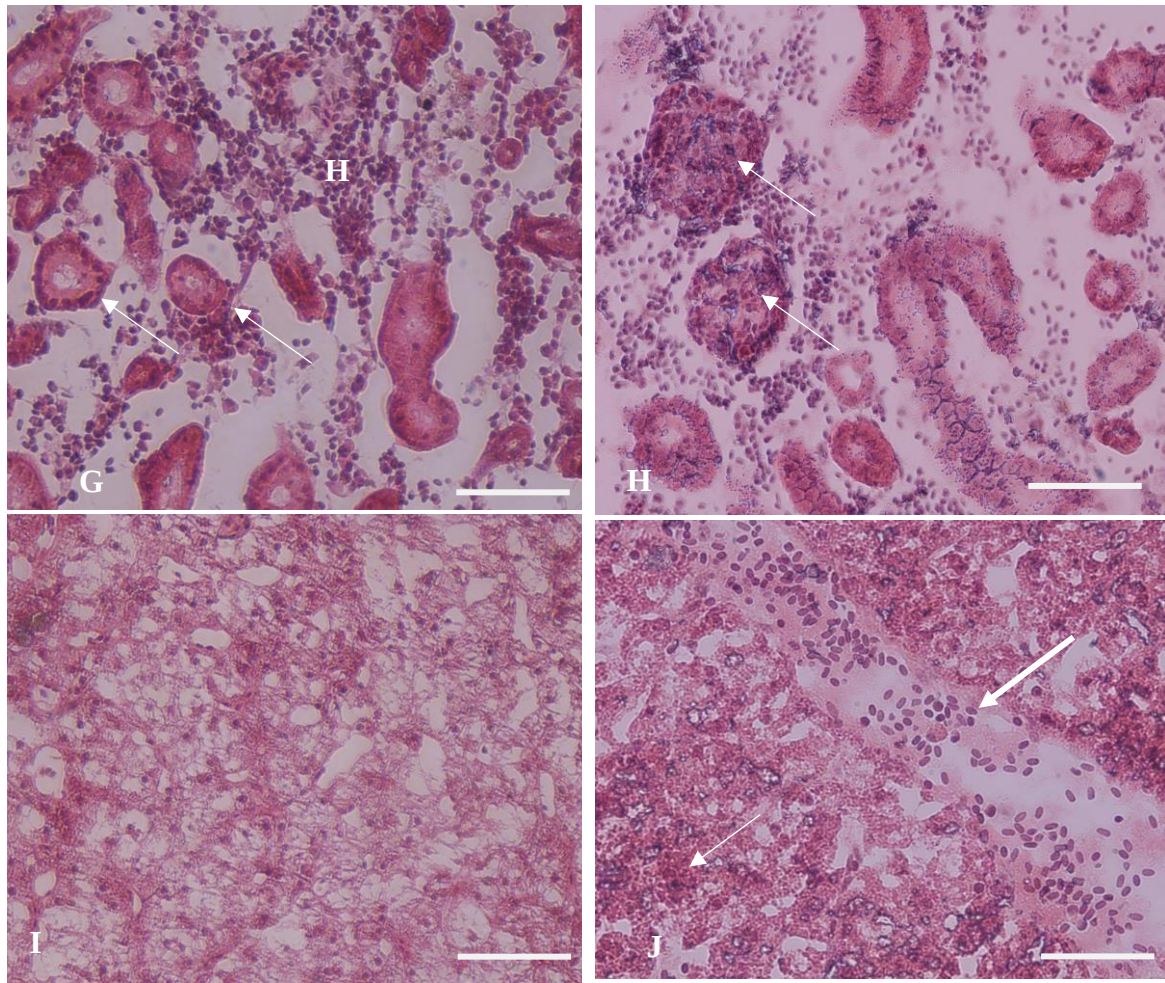


Figure 3.7: Photomicrograph of spleen, intestine, gills and kidneys of control and infected empurau. Bar = 50 μ m **(A)** Photomicrograph of control spleen showing normal architecture with ellipsoids (E) beside white and red pulps, 200X, H&E **(B)** Severe splenitis with diffuse fibrinoid necrosis of ellipsoidal sheath (white arrows) at 200X, H&E **(C)** Photomicrograph of control intestine showing its characteristics layer: lumen, mucosa, submucosa and serosa, 100X, H&E **(D)** Intestine of infected empurau showed epithelial necrosis with sloughed necrotic debris in its lumen (showed with arrows, 200X, H&E **(E)** Photomicrograph of control gills showed normal primary lamellae (PL) and secondary lamellae (SL), 200X, H&E **(F)** Infected gills showed thickening and clubbing of secondary lamellae (arrows), 200X, H&E **(G)** Photomicrograph of control kidneys with many renal tubes (arrows) and an area of hematopoietic tissues in between [H] 200X, H&E **(H)** Infected kidney showing renal tubular necrosis and interstitial congestion (arrows), 200X, H&E **(I)** Photomicrograph of control liver showing the normal organization of polygonal hepatic cells and blood capillary, 200X, H&E **(J)** Infected liver showed focal area of lymphocytic aggregation (thin arrow) migrated from dilated blood vessels (thick arrow) towards hepatic cells, 200X, H&E

3.3.5 Drug Sensitivity Test

Table 3.4 displays the drug sensitivity test results of CM-01 strain, serving as an extension of a past study characterizing its susceptibility to antibiotics (Kho et al., 2023). Antibiotics tested susceptible to the strain includes ceftazidime, cefepime, ceftolozane/tazobactam, piperacillin/tazobactam, ceftazidime/avibactam, ceftolozane/tazobactam, doripenem, imipenem, meropenem, and ciprofloxacin. Furthermore, resistance was found towards four antibiotics, including amoxicillin, cefotaxime, cefadroxil, and ertapenem. The strain was concluded to inhabit an environment subjected to regular antibiotics administration due to its MAR index (0.44), which further aligns well with the previous study (Kho et al., 2023).

Table 3.4: Antibiotics sensitivity of CM-01 to different drugs

	Antimicrobial agents	Concentration (µg/mL)	CM-01
Penicillin	Amoxicillin	2-32	Resistant
Cephalosporin	Ceftazidime	1-32	Susceptible
	Cefotaxime	1-64	Resistant
	Cefadroxil	1-32	Resistant
	Cefepime	1-64	Susceptible
β-lactam combination agents	Piperacillin/Tazobactam	4/4-128/4	Susceptible
	Ceftazidime/Avibactam	4/4-128/4	Susceptible
	Ceftolozane/Tazobactam	4/4-128/4	Susceptible
Carbapenems	Doripenem	0.25-16	Susceptible
	Ertapenem	0.25-16	Resistant
	Imipenem	0.25-16	Susceptible
	Meropenem	0.25-16	Susceptible
Fluoroquinolones	Ciprofloxacin	0.25-4	Susceptible

3.3.6 Virulence Genes

In this study, all twelve virulent genes were screened, further revealing their presence within the genome of CM-01 strain (Figure 3.8).

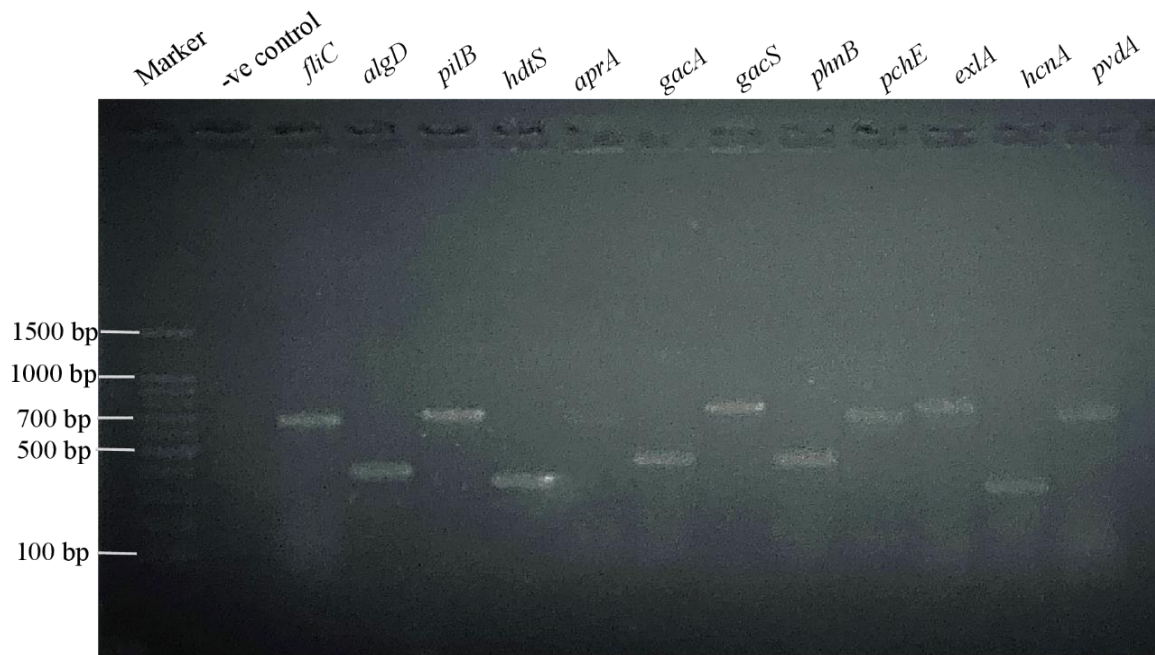


Figure 3.8: Agarose gel electrophoresis visualizing the amplification products of virulence genes. Twelve virulent genes (*fliC*, *algD*, *pilB*, *aprA*, *gacA*, *gacS*, *pchE*, *phnB*, *hcnA*, *hdtS*, *exlA* and *pvdA*) were screened and found present within CM-01 with the size estimated through 100 bp DNA ladder (Transgene, French)

3.3.7 Growing Characteristics

Figure 3.9 shows the growth characteristics of the strain under various conditions. The figure illustrates that the growth and concentration of the isolate were optimized at pH 7 (Figure 3.9A). Furthermore, a similar growth pattern was observed under pH 5 and 7 except for an extended latent phase observed in the growth curve under pH 5. Growth was completely inhibited under pH 3, 9, 10, and 11. Regarding temperature, growth and concentration exhibited similar patterns at 25 °C and 30 °C (Figure 3.9B), although a longer latent phase was observed at 25 °C. There was completely no growth at 42 °C.

The growth curve depicted under 30 °C and pH 7 (Figure 3.9C) illustrates that the latent phase occurred from 0 to 12 hours, followed by the logarithmic phase from 12 to 20 hours, with the stationary phase recorded between 20 to 24 hours. Later from 24 hours onwards to 40 hours, a slower growth curve termed secondary growth phase was observed. Subsequently, the strain entered the aging phase in 40 hours as the concentration and growth of the strain went downwards.

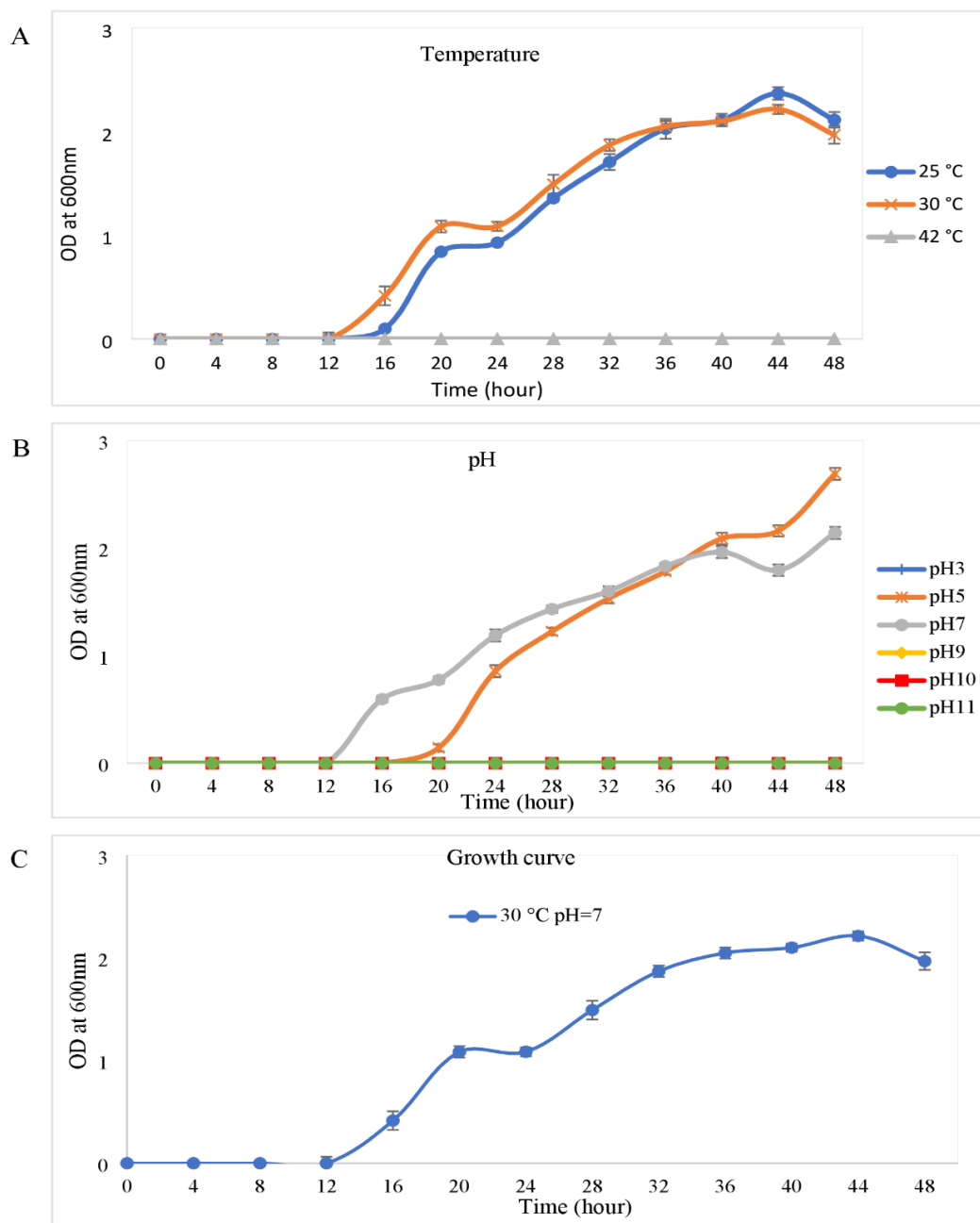


Figure 3.9: Growing characteristics of *Pseudomonas koreensis*, CM-01 strain. **(A)** Growth at temperature 25 °C to 42 °C **(B)** and pH 3 to 11. **(C)** The growth curve of CM-01 at pH 7.0 and 30 °C exhibited four phases of bacterial growth

3.4 Discussion

The genus *Pseudomonas* is widely distributed across various environments, including air, water, soil, and sediments, and can infect diverse host species, such as animals, plants, algae, and fungi (Govarthanan et al., 2019). Many *Pseudomonas* members are commonly found in agricultural soils due to their roles in organic matter decomposition, promoting plant growth, and sometimes causing pathogenic effects (Palleroni, 1993). Previous studies have highlighted several ways in which *Pseudomonas* can impact the environment, fish, and ultimately humans as consumers:

- i) *Pseudomonas aeruginosa* can serve as an indicator of low water quality (Mena & Gerba, 2009).
- ii) Identification of *Pseudomonas* species as freshwater fish pathogens, sometimes causing no symptoms but also associated with haemorrhage and eye lesions.
- iii) The extensive utilization of antibiotics in aquaculture farms can contribute to the spreading of antibiotics-resistant *Pseudomonas* species.
- iv) There is potential for *Pseudomonas* species to be used as biocontrol or anti-pathogen agents against fungal infections (Liu et al., 2015; Austin & Austin, 2016; Duman et al., 2021).

Pseudomonas koreensis falls under *Pseudomonas fluorescens* complex and can be found across a wide range of environments (Garrido-Sanz et al., 2016). *P. koreensis* was firstly isolated from Korean agricultural soils back in early 2000's (Kwon et al., 2003). In addition, it can be found in freshwater fishes (Shahi & Malik, 2014), rhizospheres (Lozano et al., 2017), agricultural soils (Lopes et al., 2018), yak milk (Kaur et al., 2019) and patients with bronchiectasis (Fluit et al., 2020).

3.4.1 Characterisation of *P. koreensis* CM-01

Morphological characteristics are indeed valuable for the taxonomic differentiation of *Pseudomonas* species (Palleroni, 1983). *P. koreensis* typically presents as Gram-negative, oxidative and non-fermenting bacilli, a description consistent with findings from a study of 90 *Pseudomonas* strains isolated from a fish farm in Turkey (Duman et al., 2021). This morphology is reflected in its growth on Eosin-Methylene Blue (EMB) agar, producing colorless colonies, as well as on MacConkey agar, where lactose fermentation is not observed (Figure 3.1). In the API 20E system, *P. koreensis* exhibits negative results for lactose fermentation and utilization of nine other sugars, including glucose, mannose, inositol, sorbitol, rhamnose, sucrose, melibiose, amygdalin, and arabinose. Nevertheless, it had shown the ability to utilize citrate and arginine as carbon and energy sources, as evidenced by positive results in the citrate test (Koser, 1923). For bacterial identification, 16S rRNA sequencing is a powerful and commonly used method (Busse et al., 1996) (Figure 3.3). In this study, construction of phylogenetic tree was performed via the 16S rRNA sequence, which displayed a monophyletic relationship of CM-01 strain with *P. koreensis* BS3658, further confirming its taxonomic classification (Figure 3.4).

The growth of bacterial colonies under various pH and temperature conditions can display a unique pattern of turbidity, influenced by the optimal temperature and pH for bacterial enzyme production. This aspect holds significance for controlling pathogens in the agriculture and aquaculture sectors (Alexander et al., 1978). Research by Christen & Marshall (1984) has highlighted the impact of temperature and pH on the growth of *Pseudomonas* species. Biological control implementation measure is often a safer option as compared to chemical control, despite being equally efficient and reliable (Powell, 1993). In the case of *P. koreensis* CM-01, growth was observed at temperatures of 25 °C and 30 °C,

as well as pH levels of 5 and 7 (Figure 3.9). Notably, at pH 5, a slightly prolonged latent phase was observed compared to pH 7, mirroring a similar trend seen at 25 °C compared to 30 °C. These variations could be attributed to the suboptimal temperature and pH conditions, inducing environmental stress. During the lag phase, bacterial cells undergo adaptations necessary to take new advantage of the new environmental conditions (Madigan et al., 2000), which may involve the synthesis of cellular components essential for growth (Rolfe et al., 2011). The findings regarding optimal growth conditions and biochemical characteristics of the *P. koreensis* isolate provide a solid foundation for further investigation into this pathogen.

3.4.2 Experimental Challenge

For LD₅₀ value falls between 10⁶ to 10⁷ CFU/g of fish body weight, the pathogen can be categorized as moderately virulent (Mittal et al., 1980) (Figure 3.5). The LD₅₀ of the bacteria of interest of this study is calculated as 5.01x10⁷ CFU/mL via Miller and Tainter method, indicating its moderate virulence to *Tor tambroides*. A significant mortality rate was seen within the first 24 hours after the challenge, with a subsequent gradual increment until zero mortality occurred beyond the experimental challenge period (Table 3.3). There were no deaths or clinical symptoms observed within the control group. CM-01 strain was suggested to own a moderately brief incubation period and disease advancement while in comparison to other studies with high mortality rate occurred within the first 12 hours (Chen et al., 2019a; Han et al., 2017). Another study reported deaths on the 7th day of a 15-day experimental challenge period in rainbow trout (Oh et al., 2017). These differences in disease progression may be attributed to variations in bacterial strains.

Pseudomonas species are opportunistic Gram-negative bacteria found as a part of normal gut microflora and freely across aquatic environments, which cause infection when the environmental condition changes (Angelini & Seigneur, 1988). *Pseudomonas* septicaemia is characterised by petechial haemorrhage, detached scales, abdominal ascites, exophthalmia and darkness of the skin. The causative agents of the disease were identified across wide species of *Pseudomonas*, includes *P. fluorescens*, *P. aeruginosa*, *P. anguilliseptica* and *P. putida* (Austin et al., 2007). Different *Pseudomonas* species tend to cause similar clinical symptoms, but sometimes there will be dissimilarities observed. Besides, the same bacteria can even cause divergent clinical symptoms in different fish (Chen et al., 2019a). Past isolation of *P. koreensis* from *Tor putitora* had shown eye opaqueness, haemorrhage in the periphery of the eyes and internally pale liver (Shahi & Malik, 2014). In *Tor tambroides* infected with the CM-01 isolate, notable clinical symptoms such as internal fluid accumulation, loss of scales, coelomic distension, petechial haemorrhage, pop-eyes, and inflamed vent were observed (Figure 3.6). Nevertheless, these symptoms were not shown following infection with *P. koreensis* in the study reported by Shahi & Malik (2014) where opaqueness of eyes was not manifested in the infection of this study but can be further explained through the difficulty of reproduction of similar results using *in vivo* bacterial challenge experiments.

Histopathological observations revealed significant changes in liver, gills, kidney, intestine and spleen caused by the CM-01 isolate (Figure 3.7). Unlike a previous study performing experimental challenges using the same bacteria on *Tor putitora* (Shahi & Malik, 2014), which primarily affected the spleen, severe splenitis was observed alongside epithelial necrosis in the intestine in the current study. Such necrosis can be explained by the accumulation of bacterial cells within the intestinal lumen (Abdelhamed et al., 2017). As

shown by the histopathological observations within gills in *Ictalurus punctatus* infected with *Aeromonas hydrophila* (Abdelhamed et al., 2017), there were thickening and clubbing observed within secondary lamellae as well. Additionally, the infected kidney showed renal tubular necrosis and interstitial congestion, while the infected liver displayed focal areas of lymphocytic aggregation migrated from dilated blood vessels. These symptoms suggested the development of acute haemorrhage and necrosis within the kidney and liver. It can be attributed to the secretion of toxins and extracellular substances including hemolysin, protease, and elastase, which further results in organ failure and swift mortality. The existence of *aprA* gene responsible for alkaline protease encode further corroborates such occurrence, despite the absence of elastase within the genome of CM-01 (Abdelhamed et al., 2017). Similar histopathological changes have been observed in *Oreochromis* spp. infected with *Aeromonas hydrophila* (Hamid et al., 2017), *Carassius auratus gibelio* infected with *Aeromonas veronii* (Chen et al., 2019a), and *Ictalurus punctatus* infected with *Aeromonas hydrophila* (Abdelhamed et al., 2017).

The findings from the infection experiment indicate that *P. koreensis* might function as an opportunistic pathogen, particularly in situations of host stress or adverse conditions. Furthermore, research by Duman et al. (2021) revealed on the lower isolation rate of *Pseudomonas* as compared to other investigated aquatic pathogens in Turkey between 2013 and 2017. This aligns with a separate study conducted on bacteria isolated from aquaculture farms in Sarawak, Malaysia, which also identified *Pseudomonas* as one of the less frequently isolated bacteria (Lihan et al., 2020). Consequently, it is proposed that relative to other species, *Pseudomonas* can assume either opportunistic or primary invasive roles in salmonid freshwater fish (Duman et al., 2021).

Reportedly, the specific factor driving the proliferation of *P. koreensis* within *T. putitora*, resulting in pathological alterations and mortality in the fish, remains unidentified (Shahi & Malik, 2014). The origin of the *P. koreensis* CM-01 strain stemmed from a diseased *T. tambroides* sourced from a local aquaculture facility. To date, this marks the second documented instance of *P. koreensis* occurrence in fish, following its initial isolation in India in 2014.

3.4.3 Virulence and Toxicity

Virulent genes play a crucial role in assessing the pathogenicity of bacteria (Senderovich et al., 2012) (Figure 3.8). In this investigation, twelve virulent genes (*fliC*, *algD*, *pilB*, *aprA*, *gacA*, *gacS*, *pchE*, *phnB*, *hcnA*, *hdtS*, *exlA*, and *pvdA*) were examined, with all were detected within *Pseudomonas koreensis* CM-01. These genes encompass various factors essential for pathogenicity, including iron uptake (*pchE*, *pvdA*), alkaline protease (*aprA*), acylhomoserine lactone synthase (*hdtS*), *gacS/gacA* component regulation system (*gacS*, *gacA*), type IV secretion system (*phnB*), hydrogen cyanide production (*hcnA*), exolysin (*exlA*), alginate biosynthesis (*algD*), flagella (*fliC*) and pili (*pilB*).

The *fliC* gene encodes flagellin, which plays a role in bacterial motility, attachment, and chemotaxis (He et al., 2012). Additionally, it has been observed to potentially function as a secretory system responsible for different virulence factor transport (Aizawa, 2001). Back in the year 2002, it was revealed as the virulent-strain-specific protein (Tan et al., 2002) which was further identified as an antigenic protein through its involvement in rabbit polyclonal antiserum (Verjan et al., 2005). Furthermore, the presence of another adherence gene, *pilB*, was identified in *P. koreensis* CM-01. The formation of pili is encoded by *PilB*, facilitating twitching motility, which enables displacement on solid surfaces through the

extension and retraction of pili, resulting in a 'swarming' behavior (Asikyan et al., 2008). This suggests that pili may serve as receptors, contributing to biofilm formation and initiating bacteriophage infections (Benie et al., 2017). Additionally, the *algD* gene was identified in *P. koreensis* CM-01. *AlgD* is responsible for producing alginate, which results in the formation of mucoid colonies. These colonies serve as a protective mechanism against antibiotics and immune responses by inhibiting complement activation, thereby reducing polymorphonuclear chemotaxis and phagocytosis (Fadhil et al., 2016; Benie et al., 2017). Alginate is involved in biofilm formation as one of the major exopolysaccharides of the biofilm matrix (Fadhil et al., 2016). The mucoid appearance can be correlated to the *algD* expression (Høiby et al., 2010).

Alkaline protease functions to degrade a variety of substances including central complement protein C3, interferon-gamma, and flagellin, is found to be encoded by *AprA* gene. It specifically degrades the monomeric form of flagellin, but not the polymeric form, rendering the degraded flagellin unrecognizable by TLR5 (Bardoel et al., 2011). TLR5 recognizes the flagellin domain essential for filament assembly and motility. The innate immune system relies on recognizing highly conserved flagellin to efficiently eliminate a broad range of microorganisms (Bardoel et al., 2011). The *GacS/GacA* system is a conserved two-component mechanism that governs virulence-related characteristics in a range of pathogens, including various species of *Pseudomonas*. Its role in bacterial virulence regulation was initially established 25 years ago through studies of *Pseudomonas syringae* (Willis et al., 1990). *GacS/GacA* negatively regulates the expression of type III secretion system (T3SS) and motility genes, while positively regulating biofilm formation and type VI secretion in *Pseudomonas aeruginosa* (Valentini et al., 2018). Furthermore, it acts as a

master regulator of multiple virulence traits, including T3SS establishment, motility, and toxin production in *P. syringae* (Chatterjee et al., 2003).

The virulence gene *PchE* has been observed to promote flagellar motility while inhibiting cell adhesion, facilitating pathogen movement to the intestinal surface during the early stages of infection. Previous research has indicated that induced or inappropriate *pchE*-dependent flagellar expression could impede cell attachment as the disease progresses (Andreozzi & Uhlich, 2020). Besides, it had been proven the regulation of *pchE* onto three adhesion-encoding genes (*fliC*, *lpfA1* and *eae*) and four master regulators (*flhDC*, *ler*, *stpA* and *csgD*) in *Escherichia coli* (Andreozzi & Uhlich, 2020). Another iron uptake regulated gene, *pvdA* gene which encodes for enzyme L-ornithine N5-oxygenase was detected within *P. koreensis* CM-01 as well. *PchE* catalyzes the pyoverdine biosynthetic pathway in *Pseudomonas aeruginosa*. Iron is essential for bacteria, although it is often unavailable under aerobic conditions at neutral pH (Neilands, 1981). Microorganisms possess high-affinity uptake systems for iron, known as siderophores (Leoni et al., 1996). Among these, pyoverdine synthesis is typically downregulated in response to the availability of iron (Rombel et al., 1995). *PhnB* gene encodes for subunits of a phenazine-specific anthranilate synthase and its mutation has been reported to reduce virulence (Mahajan-Miklos et al., 1999). Phenazines had been suggested as multi-host pathogenicity factors due to their reduced virulence in both *Arabidopsis* leaf infiltration model and mouse burn model (Mahajan-Miklos et al., 1999).

Additionally, the *hdtS* gene encodes for N-acyl-homoserine lactone synthase was identified within *Pseudomonas koreensis* CM-01. It was found to mediate N-acyl-homoserine lactone (AHL)-mediated quorum sensing, which is commonly used by various Gammaproteobacteria bacteria, including *Pseudomonas aeruginosa* (Williams et al., 2007).

The *hdtS* gene belongs to the lysophosphatidic acid acyl transferase family, also found in *Pseudomonas fluorescens* (Laue et al., 2000). Members of the *hdtS* family are involved in acylating lysophosphatidic acid and synthesizing AHL (Churchill & Chen, 2011). Moreover, *hcnA* was found to encode hydrogen cyanide (HCN) synthase along with the *hcnBC* gene. This secondary metabolite production maximizes during the transition from the exponential growth phase to the stationary growth phase as well as reduced oxygen levels and elevated cell densities (Castric et al., 1979; Castric, 1983). Evidence has shown *hcn* to act as a virulence factor in experimental infection models involving mutant *hcn* (*Pseudomonas aeruginosa*) insertions, resulting in mortality of the nematode *Caenorhabditis elegans* (Pessi & Haas, 2000). Besides, the presence of *exlA* gene which encodes for exolysin was detected within *Pseudomonas koreensis* CM-01. Exolysin is secreted by the type III secretion system in bacteria to facilitate infection establishment (Hauser, 2009; Elsen et al., 2014). Initially detected in a patient with chronic obstructive pulmonary disease (COPD) exhibiting lung hemorrhages (Elsen et al., 2014), its introduction following recombinant *exlBA* caused mortality in a mouse model infected with acute pneumonia (Reboud et al., 2017). Exolysin has been identified in various *Pseudomonas* species, including *P. putida*, *P. fluorescens*, *P. aeruginosa*, *P. protegens*, and *P. entomophila* (Basso et al., 2017). The identification of several virulence genes within *P. koreensis* CM-01 suggests potential synergistic effects arising from the combination of these genes, contributing to its pathogenicity.

3.4.4 Antibiotics Susceptibility

The drug sensitivity test on the bacteria of interest suggested possibility of antimicrobials to serve as a treatment option for *P. koreensis* infections. *Pseudomonas* species are often characterized as multi-drug resistant as they were resistant to a broad range

of antimicrobials, including ampicillin, trimethoprim+sulfamethoxazole, nitrofurantoin, aztreonam, and cefotaxime, among others (Matyar et al., 2008). Kho et al. (2023) reported on the resistance of CM-01 strain toward ampicillin, aztreonam, clindamycin, and ceftazidime. Therefore, in this study, the VITEK2 system was employed to provide further insights into the antibiotic susceptibility of the strain, with additional antimicrobials included in the analysis. As a result, the strain had shown its inability to resist ceftazidime, cefepime, doripenem, imipenem, meropenem, and ciprofloxacin. Furthermore, its susceptibility to ciprofloxacin aligns with findings from Wolska et al. (1999) and Al-Tawfiq (2007), which reported susceptibility rates of up to 99% and 85% for *P. aeruginosa* strains, respectively. However, susceptibility to ciprofloxacin appeared to decline gradually, as indicated by a 2020 study reporting a 70% susceptibility rate for isolated strains of *P. aeruginosa* (Ahmad et al., 2020).

CM-01 demonstrated resistance to amoxicillin (penicillin) (2-32 µg/mL), cefotaxime (3rd generation cephalosporin) (1-64 µg/mL), cefadroxil (1st generation cephalosporin) (1-32 µg/mL), and ertapenem (carbapenems) (0.25-16 µg/mL) (Table 3.4). Its resistance to amoxicillin aligns with findings from *P. koreensis* isolated from *T. putitora* (Shahi & Malik, 2014). *Pseudomonas* species are recognized for their resistance to various antibiotics, particularly penicillin and cephalosporins, which belong to β -lactam combination agents, often attributed to the presence of extended-spectrum beta-lactamases (ESBL) (Peymani et al., 2017). However, CM-01 exhibited susceptibility to all β -lactam combination agents tested, including piperacillin/tazobactam, ceftazidime/avibactam, and ceftolozane/tazobactam, probably because of the existence of beta-lactamase inhibitors. Therefore, it can be inferred that β -lactam combination agents might be the preferred treatment option for infections caused by *P. koreensis*. In comparison to a *P. fluorescens*

strain isolated from water in Paris, which exhibited resistance to carbapenems (Girlich et al., 2010), CM-01 displayed susceptibility to most of the carbapenems tested, except ertapenem.

There have been cases reported on patients infected with resistant multidrug-resistant (MDR) *P. aeruginosa* showing poor prognoses (Ahmad et al., 2020). Additionally, research has highlighted *P. aeruginosa* strains to have antibiotic-resistance genes (*bla*_{TEM}, *bla*_{CTX-M}, and *tetA*), further explaining their multidrug resistance against amoxicillin, cefotaxime, tetracycline, and gentamicin (Algammal et al., 2020). Furthermore, studies had warned on the frequency of isolation of antibiotics-resistant *Pseudomonas* strains from the hospital effluents (Bakon et al., 2023), likely because of the discovery of pharmaceutical residues, including antibiotics, in rivers in Selangor, Malaysia (Praveena et al., 2018). The prevalence of antibiotic-resistant strains across Sarawak, Malaysia was not left out as well. In 2016, researchers had reported on high resistance to streptomycin and ampicillin, despite claims of no history of antibiotic administration, suggesting antibiotic contamination as a significant factor (Kathleen et al., 2016). In addition, bacteria possess an ability to acquire antibiotic resistance traits via antibiotic residues left in the environment, these may include unconsumed food and the faeces of cultured organisms (Tu et al., 2009; Tamminen et al., 2011). Moreover, another antibiotic-resistant gene reservoir in farms is fish feed (Tamminen et al., 2011). In Sarawak, the emergence of MAR bacteria is being correlated with reported MAR indexes of more than 0.2 across selected farms (Lihan et al., 2020).

Therefore, it is crucial to closely monitor this issue, particularly considering the potential for *Pseudomonas* strains to infect terrestrial animals and humans. Studying the antimicrobial spectrum of the CM-01 isolate is anticipated to aid in categorizing this microorganism.

CHAPTER 4

IDENTIFICATION OF SINGLE NUCLEOTIDE POLYMORPHISMS (SNPS) ASSOCIATED WITH DISEASE RESISTANCE TOWARDS *P. KOREENSIS* IN EMPURAU (*T. TAMBROIDES*) USING T-PLEX ARMS PCR ASSAY

4.1 Introduction

Aquaculture is rapidly expanding as one of the rapidly growing food production sectors worldwide, driven by the increasing demand for fish and fish products that provide proteins and essential micronutrients for balanced nutrition and good health (FAO, 2016; DOF, 2019; Jumatli & Ismail, 2021). Fish production had achieved up to 17.3 % of the global population animals' protein intake and 6.8 % of all proteins taken in the 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 Fisheries Sector includes capture fisheries, inland fisheries and aquaculture has produced up to 5.7 million tones with an estimated value of RM4 billion with an increment of 37.5 % and 17.2 % in quantity and value, respectively as compared to the year 2021 (DOF, 2022).

Tor tambroides, or locally known as “empurau”, is one of the prized species within *Cyprinidae* family, valued not only for human consumption, but also its ornamental and sporting values. Despite the excessive fishing threats due to its unique flesh taste, the species face dangers from artificial propagation activities for both aquaculture production and conservation purposes as well (Ingram et al., 2005; Kottelat et al., 2018; Lau et al., 2022). Despite 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 cases of viral and bacterial infections reported within local fish farms, further causing huge economic losses to the industry (Chiew et al., 2019).

In fish, infectious diseases are often caused by a variety of agents such as bacteria, viruses, fungi and parasites. Among these, bacterial pathogens are the primary culprits behind most fish farming illnesses (Irshath 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 fins and gills (Lau et al., 2022; Lau et al., 2024). Besides, back in year 2014, *P. koreensis* was documented as the causative agent for eye lesions in gold mahseer (*Tor putitora*) in India (Shahi & Malik, 2014). Despite its limited disease occurrence in fish, public health focuses much of its attention on *Pseudomonas* infection as it causes serious foodborne infection due to raw fish and byproduct consumption (Zilberberg & 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 most immediate choice is the usage of antibiotics. However, the extensive and 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). Although water environment improvement can control survival rate to some extent,

enhancement in the ability to disease resistance is suggested as an 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. In the year 2019, World Health Organization (WHO) had reported on 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).

Developing disease-resistant broodstock has emerged as a promising strategy to address antimicrobial resistance (AMR) concerns in aquaculture (Moss et al., 2012). Marker-assisted selection (MAS) presents a more efficient, time-saving, and environmentally friendly alternative to traditional breeding techniques (Lu et al., 2018; Zhang et al., 2019). In this study, it is believed that pooled-sequencing (Pool-Seq) is a more cost-effective alternative to SNP identification. Several studies had 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).

A recent growth in genomic technologies has enabled molecular markers such as single nucleotide polymorphism (SNP) to be readily available. It is cheaper to run in dense panels as compared to microsatellites, thus allowing more individuals to be genotyped at a greater genome coverage level (Vignal et al., 2002). SNPs that were 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 relative 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). Numerous SNPs associated with the target traits are expected to be discovered, thus it is essential to have an approach to allow the SNP marker to be easily detected not only in laboratories but also in aquaculture settings for selective breeding among the wild populations. 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 can distinguish the amplified allele-specific amplicons in a single tube via melting temperature.

Over the years, molecular genetic studies have made great progress in facilitating selective breeding through discovery of SNPs related to economic traits. Considering the significant economic and cultural value of *Tor tambroides*, any disease occurrence could potentially lead to severe repercussions. Therefore, in this chapter, it is aim to detect and identify the candidate disease-resistant SNP markers towards *Pseudomonas koreensis* in *Tor tambroides*. This can be achieved by performing whole genome Pool-Seq on both resistant and susceptible groups of fish are aimed to be detected and identified. 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 empurau breeding.

4.2 Materials and Methods

4.2.1 Bacterial Challenge and Sample Preparation

For the bacterial challenge, 60 juvenile *Tor tambroides* were randomly distributed into 6 tanks. 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 with *P. koreensis* at 5×10^7 CFU/ml (LD₅₀) while the other group were acted as a control group. After introduction of *P. koreensis* into the fishes via peritoneal injection, all fishes were raised for 7 days at 25 °C for observation. All the experimentally infected fishes were observed at hourly interval for the first 12 hours and at 6 hours interval for the subsequent observation period. Moribund or dead fishes observed during the post-infection challenge were categorized as susceptible group (SG) while those fish survived the challenge were considered as resistant group (RG). Twenty samples from SG and 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 a modified CTAB-based extraction method.

4.2.2 Sequencing and Read Cleaning

In this study, pooled sequencing approach was implemented which provides an alternative and cost-effective way to sequence DNA (Schlötterer et al., 2014). A total of 20 individuals were made up of both resistant and susceptible pools with each respective DNA pooled together at a concentration of 40 µg/µL, with each DNA sample of 2 µg/µL. Approximately one 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 before proceeding with SNP calling (Chen et al., 2018).

4.2.3 Reads mapping, SNP calling and Identification

Trimmed paired-end reads of both resistant and susceptible reads were mapped to *Tor tambroides* reference genome (GCA_021397915.1) (Lau et al., 2022). The reference genome was prepared before sequence alignment using Burrows-Wheeler aligner (BWA) *index* command (Li & Durbin, 2010). The reads were mapped using the BWA *mem* command. Samtools version 1.13 (Danecek et al., 2021) was then used for sorting the BAM file and subsequently create a 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 PoPoolation2 script mpileup2sync.jar by keeping base quality of 20. Next, SNPs' allelic frequency (script: snp-frequency-diff.pl) was used to maintain a minimum allele count of 6, minimum and maximum coverage of 50 and 200, respectively. To investigate the presence of any consistent allele frequency changes belonging to two sample groups, Cochran-Mantel-Haenszel (CMH) test 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 same sync file was used to determine the significance of per nucleotide allele frequency differences between two groups 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).

4.2.4 T-plex real-time PCR Assay

A total of 5 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 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 & 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).

For SNPs genotyping, T-plex ARMS real-time PCR assay had been 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) (Figure 4.1 A&B).

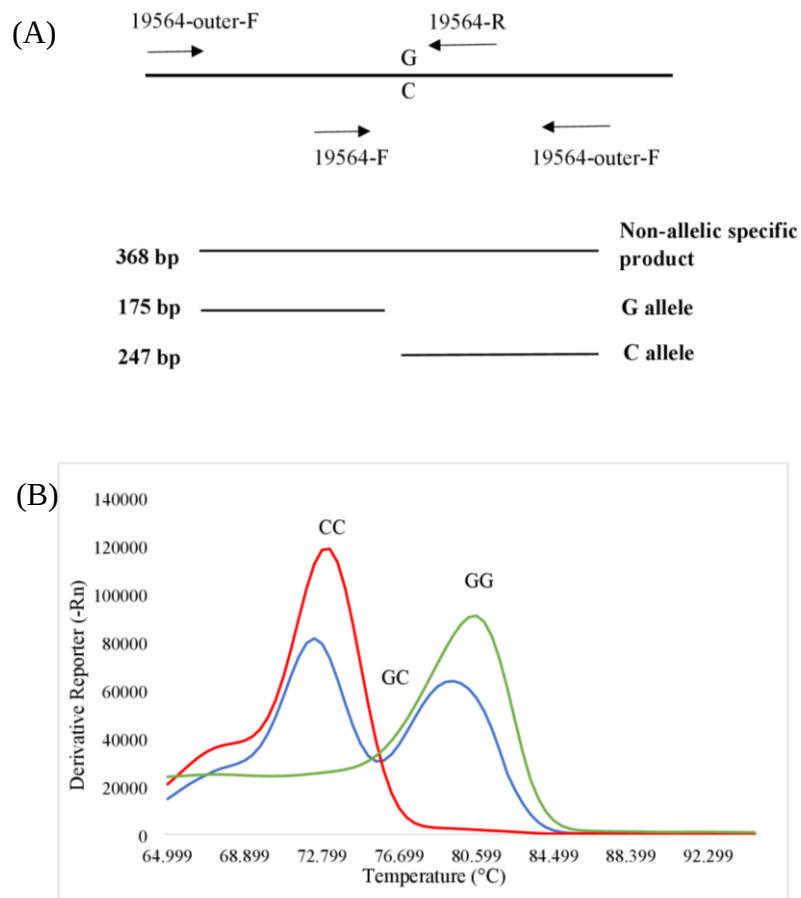


Figure 4.1: Introduction of T-Plex real-time PCR assay **(A)** It shows a schematic diagram of tetra-primer ARMS PCR for 19564 G/C genotyping in this study as an example. During the PCR amplification, a 368 bp amplicon were amplified by the non-allelic outer primer, and the allele-specific amplification 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 19564 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

4.2.5 Primer Design

The T-Plex real-time PCR assay was used to genotype the 6151C/T, 112388 G/A, 59747G/A, 27857A/T and 19564G/C. Primers were designed based on the sequences available on the public databases (SRX21910145 and SRX21910146) using web-based software (<http://primer1.soton.ac.uk/primer1.html>) (Ye et al., 2001; Collins & Ye, 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 4.1.

Table 4.1: Primer sequences for potential SNP validation used in this study were designed using web-based software. 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

SNP (Contig)		Sequences	Expected size (bp)	SNP location	SNP type
6151 C/T (WUC4741 4__np12)	F	CTCCACCACAGAAGAAGTG CTG	413	6151	C/T
	R	ACATGAACAGACATCTGAG TGAGTG	-		
	C	GTCACCTCTTCACCGGATAAA CAAAC	246		
	T	AGGGACAGATGAGACACCC A	184		
112388 G/A (WSC241__ np12)	F	GTTCTTGCGGCTGGACCGGC	364	112388	G/A
	R	GAACGACCACACGTGGTAC ATCCTG	-		
	A	GGTTTGCGTTCCCCTGGTGT CA	246		
	G	AAACGGGCACTGTGCCTCTC C	156		
59747 G/A (P_RNA_sc affold_1984 __np12)	F	GCTTTGGGTGTACGGCAAT AC	387	59747	G/A
	R	CACATACTTTTGAAAATCCA GCGTTTAG	-		
	G	CTAAACGTCTGATGCACAG GG	247		
	A	GTGTTTCCACTGTTGTGGGT	181		
27857 (WSC6179_ __np12)	F	TTTTGAGTTATTTCCGATCG CTCC	358	27857	A/T
	R	CAAAACACCTGCTAACTCC AACC	-		
	A	TCGAGAGTTGCTGACTGGA TTTT T A	247		
	T	CTTGTCTCAACACTATGAGT CAACAA	161		

Table 4.1: continued

19564 (WSC72258 —np12)	F	GGATGTTTTCAACTGAGGA GCAG	368	19564	G/C
	R	GGAGTTGTGTAGCTCACTA ACAATG	-		
	G	TGACAGATGTAGTGAAATG TCTCTCAAG	175		
	C	GAGCAGTGGATTCAAAATT TCAGAATCTG	247		

4.2.6 SNP Validation

Real-time PCR was conducted using Thunderbird Next SYBR qPCR Mix (Toyobo; Osaka, Japan) and QuantStudio 5 (Applied Biosystem; 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), 1 μ L of genomic DNA dilution (10 ng/ μ L) 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 °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 QuantStudioTM Design & Analysis Software version 1.5.3. For verification of the genotypes identification results of qPCR assay, 6 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 were visualized and analyzed using QSV analyzer (Carr et al., 2009).

The raw whole genome pooled sequencing data of both RG and SG are publicly available under NCBI Bioproject PRJNA1021700.

4.3 Results

4.3.1 General Overview and SNP Identification

A total of 50.87 Gb and 50.48 Gb of clean data was obtained from the RG and SG sequencing libraries respectively (Table 4.2). Both sample groups showed a high mapping rate 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 subjected to sync file generation prior to PoPoolation2 analysis. A total of 1,048,576 SNPs had been successfully called from the sync file generated from both groups.

Table 4.2: Summary statistics for both resistant group and 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	

Among them, 629,670 SNPs were putative transitions (including 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) (Figure 4.2). The distribution of transition SNPs was higher than transversion SNPs by two-fold. G/A and C/T were the most commonly observed transition SNPs, recorded at 173343 and 173903, respectively. As for the transversion SNPs, there were 64512 A/T and 63510 T/A SNPs observed.

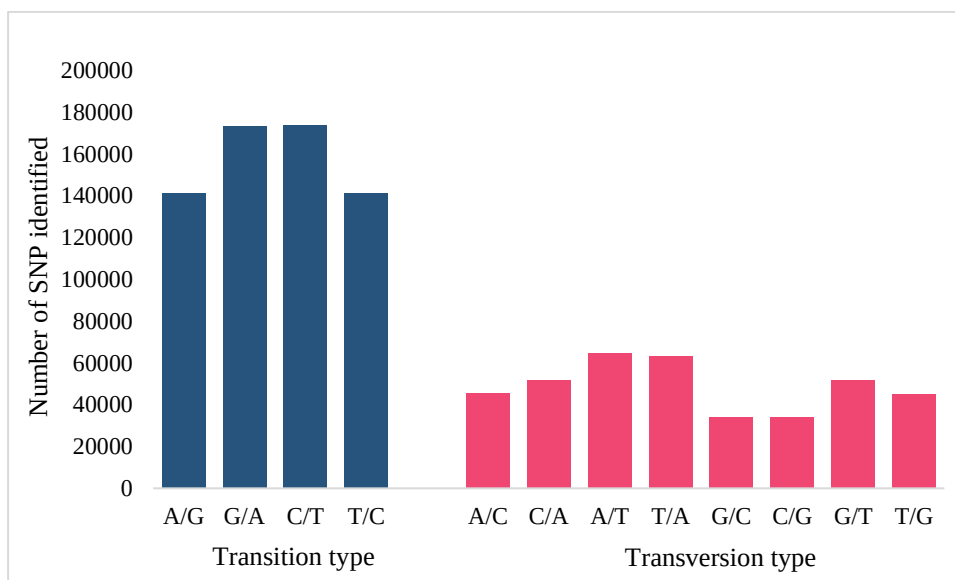


Figure 4.2: Classification of SNP identification into transition and transversion types identified within the genome of *T. tambroides*

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 top 1% F_{ST} values and p -value from FE test along with p -value < 0.001 in CMH test. Nucleotide distribution among both groups was compared for the subsequent validation step (Table 4.3). Table 4.3 shows the selected SNP markers along with their respective contig and parameters, including 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.

Table 4.3: SNP selection 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)	p -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}
112388 G/A	0:0:0:70:0:0	27:0:0:50:0:0	3.51×10^{-15}	0.2136	1.92×10^{-9}
59747 G/A	0:0:0:51:0:0	28:0:0:32:0:0	6.80×10^{-14}	0.3064	8.03×10^{-10}
27857 A/T	0:113:0:0:0:0	23:76:0:0:0:1	6.38×10^{-14}	0.1314	5.33×10^{-9}
19654 G/C	0:0:0:114:0:0	0:0:27:97:0:0	2.42×10^{-13}	0.1228	5.31×10^{-9}

4.3.2 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 4.4). 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 of 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 be having 90.6% similarity to E3 ubiquitin protein ligase TRIM35-like.

The second SNP marker, 112388 G/A, falls within the coding region which encodes for adrenoceptor alpha 2C (97.23 %). NCBI conserved motif and Interproscan had reported on the presence of alpha2C subtype adrenoceptor. It is associated with cGMP-PKG signalling and neuroactive ligand-receptor interaction pathways in KEGG (Table 4.4). As for COG databases, its categorization mainly falls within information storage and processing, including transcription, signal transduction mechanism, replication, recombination and repair.

Subsequently, 59747 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 mainly focusing on signalling and cellular processes (enzymes, exosomes and cytoskeleton proteins) by both KEGG and COG databases.

The fourth SNP marker, 27857 A/T falls as an intron which was characterised 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 categorised as cell adhesion molecules (Table 4.4). 27857 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 COG database.

Lastly, 19564 G/C SNP fall 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 has an association with ubiquitin carboxyl-terminal hydrolase and ATP-dependent DNA helicase (Table 4.4). Within KEGG BRITE, it was mainly classified as metabolism and genetic information processing, with 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 19564 G/C SNP to fall under posttranslational modification, recombination and repair as well as replication, recombination and repair category.

Table 4.4: Functional annotations of the five selected SNP markers. 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

SNP (location)	COG	KEGG BRITE (Gene and Proteins)	KEGG Pathway	KEGG Ortholog
6151 C/T (intron within gene 1)	Posttranslational modification, protein turnover, chaperones	Enzymes	RNA Polymerase	RPB1
		Transcription machinery	Nucleotide excision repair	TRIM35
		DNA repair and recombination proteins	Huntington disease	TRIM39
		Ubiquitin system		
		RNA polymerase		
112388 G/A (coding region gene 7)	at Transcription Replication, recombination and repair Signal transduction mechanism Function unknown	Peptidase and inhibitors	cGMP-PKG signalling pathway	GPR26
		Enzymes	neuroactive ligand-receptor interaction	CPN1
		Transcription factors	type I diabetes mellitus	NKX5
		G protein-coupled receptors		ADRA2A, ADRA2B, ADRA2C
				CPE, CPD, CPZ AEPB1

Table 4.4: continued

59747 G/A (intron within gene 1)	Cytoskeleton	Enzymes	cGMP-PKG signalling pathway	FLNA
		Exosomes	focal adhesion	PARP7S
		Cytoskeleton proteins	Salmonella infection	
			proteoglycans in cancer	
27857 A/T (intron within gene 2)	Replication, recombination and repair	Membrane trafficking	Cell adhesion molecules	NRCAM
	Signal transduction mechanism	Cell adhesion molecules		CNTN2, CNTN3, CNTN4, CNTN5
		Glycosylphosphatidylinositol (GPI)-anchored proteins		
19654 G/C (intergenic sequences)	Replication, recombination and repair	Enzymes	-	USP22_27_51
	Posttranslational modification, protein turnover, chaperones	Peptidase and inhibitors		USP35_38
	Function unknown	Mitochondrial biogenesis		USP_26_37
		Transcription machinery		P1F1
		Chromosome and associated proteins		
		Ubiquitin system		
		DNA replication proteins		

RPB1: DNA-directed RNA polymerase II subunit RPB1; TRIM35: tripartite motif-containing protein 35; TRIM39: tripartite motif-containing protein 39; GPR26: G protein-coupled receptor 26; CPN1: carboxypeptidase N catalytic subunit; NKX5: homeobox protein Nkx-5; ADRA2A: adrenergic receptor alpha-2A; ADRA2B: adrenergic receptor alpha-2B; ADRA2C: adrenergic receptor alpha-2C; CPE: carboxypeptidase E; CPD: carboxypeptidase D; CPE: carboxypeptidase Z; FLNA: 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

4.3.3 SNP Validation via T-plex Real-time PCR Assay

Twenty samples in SG and twenty samples in RG were used to confirm the genotype of the candidate SNP markers. A total of 5 SNP markers were selected to verify and four SNPs were successfully verified (Figure S2, S3, S4, S5, S6). The significantly different alleles from the SNP markers were detected between both RG and SG using X^2 test with one SNP marker successfully showed significant differences between both groups (Table 4.5). Out of the verified SNP markers, only 6151 C/T SNP showed one genotype for both RG and SG, which is CC genotype. Furthermore, the same genotype distribution was observed for both 112388 G/A and 27857 A/T SNP, with most of the individuals showing similar homozygous genotypes in both groups, eventually bringing insignificant statistical differences in the particular SNP for both populations. 19564 G/C SNP shows greater GG genotype distributions in RG while in SG, greater CC genotype distributions were observed. The T_m graph of each SNP marker for both RG and SG was summarized in Figures S1, S2, S3, S4 and S5.

Table 4.5: 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	
112388 G/A	GG	20	17	3.24 (0.072)
	AG	0	3	
27857 A/T	TT	20	17	3.24 (0.072)
	AT	0	3	
19564 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, 112388 G/A, 27857 A/T and 19564 G/C) had portrayed that the SNP genotypes are consistent with the grouping results of T-plex real-time PCR analysis (Figure 4.3) with red-colored line graph represents RG and blue-colored line graph represents SG. For 6151 C/T SNP, selected RG samples (R16, R17, R20) and SG samples (S7, S10, S12) had shown T_m of around 76 °C (Figure 4.3B) which can be further supported by the multiple alignment of Sanger sequencing results of showing that the T_m belongs to C allele (Figure 4.3A). 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 (Figure 4.3C & D).

As for 27857 A/T SNP, all selected samples from RG (R1, R12, R15) and SG (S1, S2, S12) had shown single T_m peak at 74 °C which is supported by Sanger sequencing results (Figure 4.3E & 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, S2 and S20 showing single peaks at different T_m of 79 °C and 72 °C, respectively (Figure 4.3G & H).

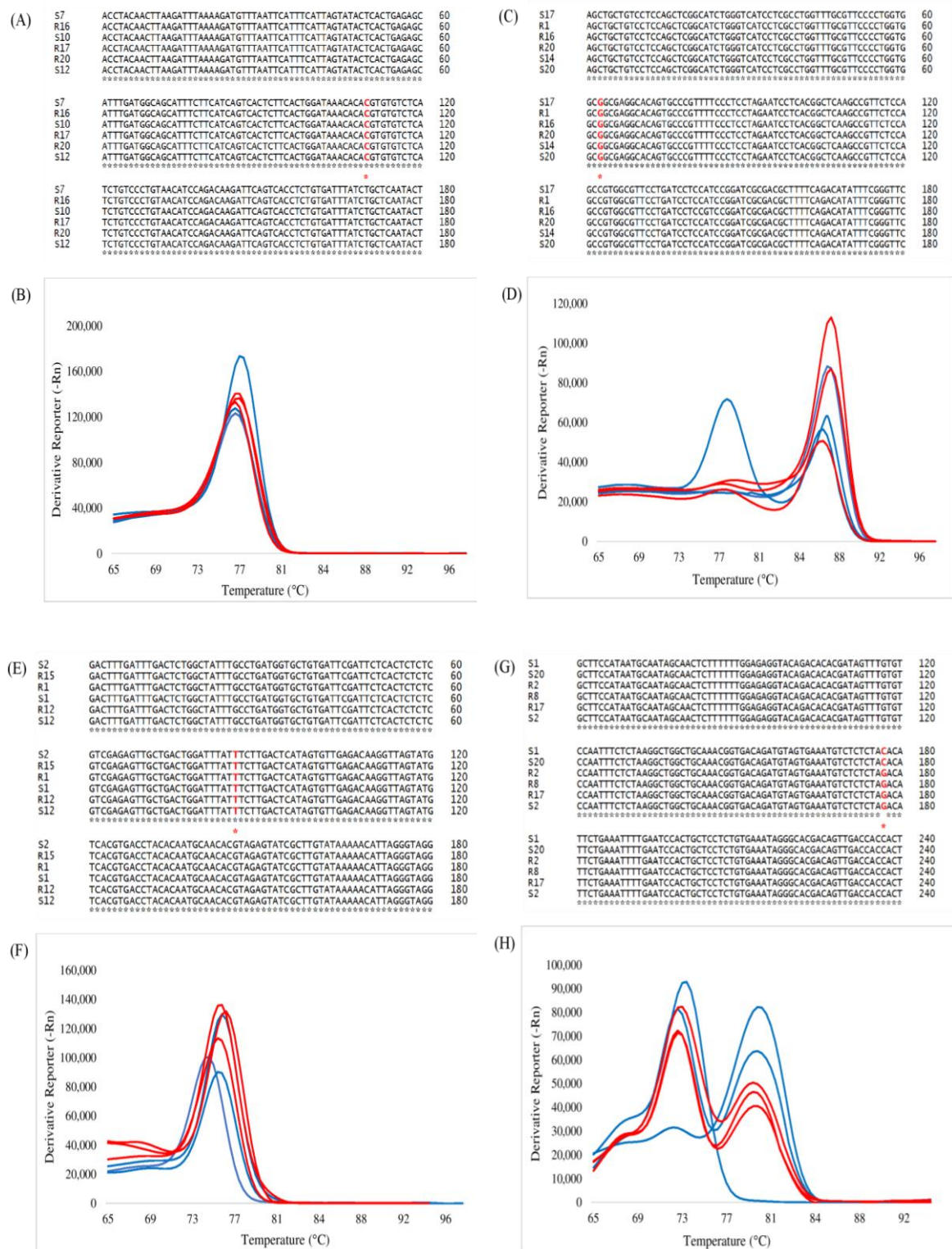


Figure 4.3: The results of Sanger sequencing and T-plex real-time PCR analysis on 6151 C/T, 112388 G/A, 27857 A/T and 19564 G/C SNP. **(A)** Multiple sequence alignment of PCR products of 6151 C/T SNP for six samples. **(B)** Discrimination of 6151 C/T SNP marker among six corresponding samples **(C)** Multiple sequence alignment of PCR products of 112388 G/A SNP for six samples. **(D)** Discrimination of 112388 G/A SNP marker among

six corresponding samples. **(E)** Multiple sequence alignment of PCR products of 27857 A/T SNP for six samples. **(F)** Discrimination of 27857 A/T SNP marker among six corresponding samples. **(G)** Multiple sequence alignment of PCR products of 19564 G/C SNP for six samples. **(H)** Discrimination of 19564 G/C SNP marker among six corresponding samples. Red star represents SNP. RG represents resistance group (red color line graph) while SG represents the susceptible group (blue color line graph)

With the successful genotyping of 19564 G/C SNP, Figure 4.4 had summarized on 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 showed GC genotype due to the double nucleotide peak detected from the Sanger sequencing chromatogram in QSV analyser. 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 on its GC genotype. Furthermore, samples S2 and S20 portrayed single peaks with different nucleotides and T_m , with allele G at 79 °C and allele C at 72 °C, respectively. Thus, the genotype for 19564 G/C SNP can be deduced via the T_m shown by each sample and is applied for all the SNP markers (Figure S5) (Figure 4.4).

Figure 4.4: 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 19564 G/C SNP

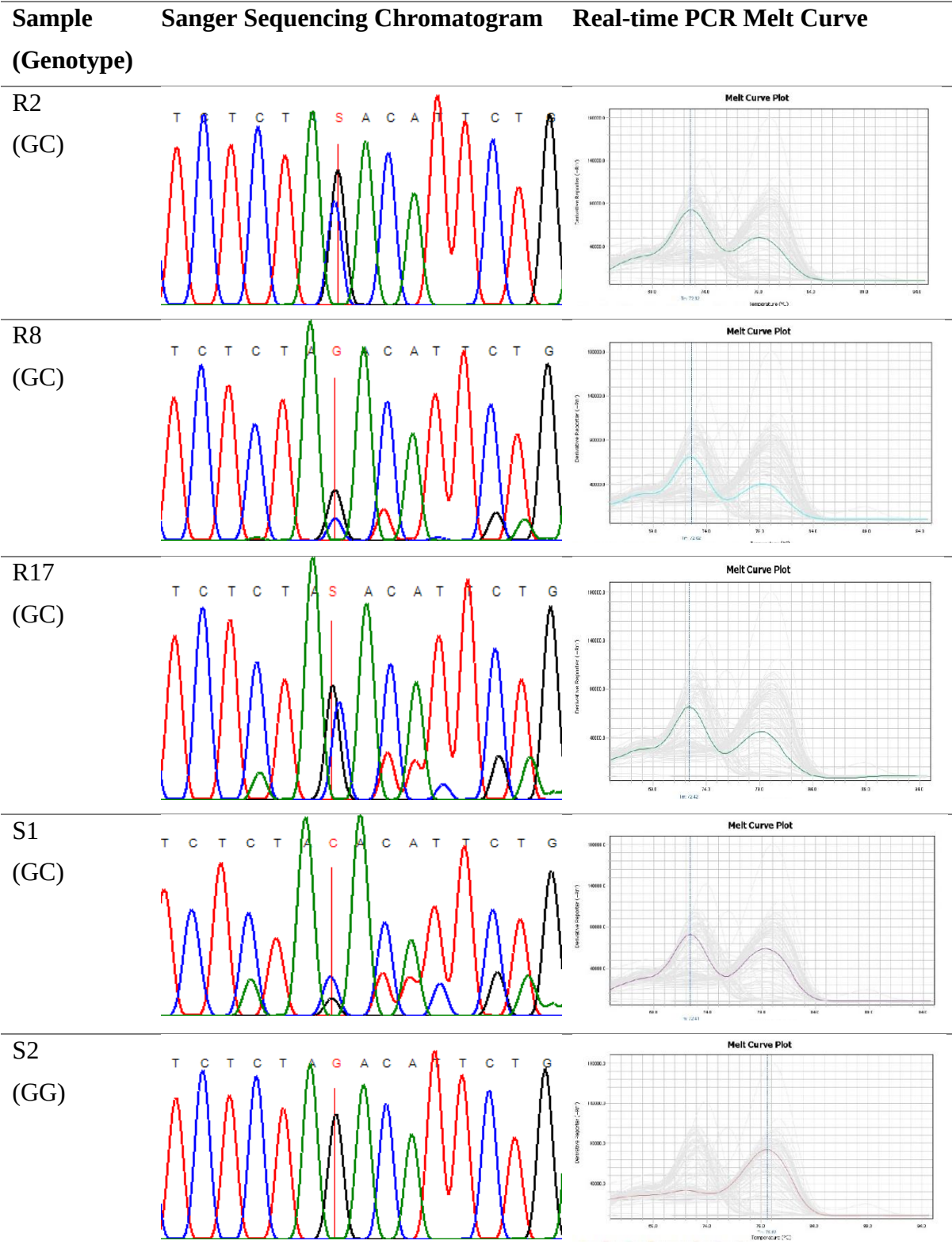
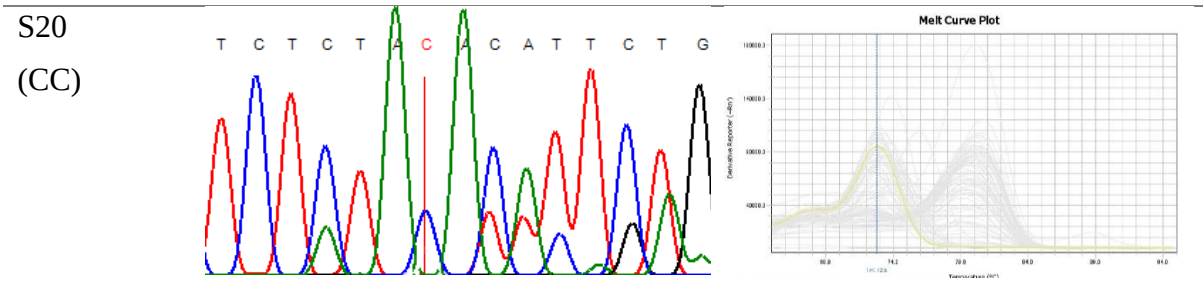


Figure 4.4 : continued



4.4 Discussion

4.4.1 Marker-Assisted Selection as Antimicrobial Resistance Solution

Bacterial infection serve as one of the significant obstacles faced by the aquaculture industry which may cause huge economic losses. Nowadays, the importance of environmental conservation has become increasingly highlighted with public consciousness, and striking a balance between production and ecological safety is crucial. The common solution to these disease outbreaks caused by microbes are the administration of antibiotics.

For decades, antibiotics had been employed as medicinal and preventative measures not only for human usage, but also applies to agricultural and aquaculture settings as well (Gajdács & Albericio, 2019). The development of antimicrobial resistance (AMR) is indeed an unavoidable phenomenon due to the nature of microbes to develop genetic mutations which further alleviate the lethal effect of antibiotics. AMR, or antimicrobial resistance, refers to the capacity of microorganisms to withstand the effects of antibiotics to which they were once susceptible. This resilience enables them to survive and proliferate despite antibiotic exposure (Zaman et al., 2017). However, there are still numerous factors that is able to contribute to such complex origins of antibiotic resistance, including inaccurate diagnosis and inappropriate prescribing practices, antibiotics overuse and misuse, inadequate healthcare facilities, reduced patient sensitivity and self-medication, extensive agricultural practices and lack of personal hygiene (Uddin et al., 2021).

In aquaculture and agricultural settings, the routine antibiotics administration for the purpose of disease prevention and growth promotion has led to incomplete antibiotic metabolism (Manyi-Loh et al., 2018). As reported by Food and Agriculture Organization of the United Nations (FAO) in the year 2012, oxytetracycline, florfenicol and trimethoprim/sulfadiazine were documented as the most commonly used antibiotics for

disease control in farms (Alday-Sanz et al., 2012). As a result, these waste products were released into the environment and further encouraged the formation of antibiotics reservoirs as well as antibiotics-resistant bacteria along with their resistance genes (Manyi-Loh et al., 2018). Consequently, human society had to face the drawbacks as diseases including pneumonia, septicaemia, tuberculosis, salmonellosis and gonorrhoea are becoming difficult to cure due to AMR (Bondad-Reantaso et al., 2023). World's Health Organization (WHO)'s latest Global Antimicrobial Surveillance System (GLASS) has reported extensive AMR across 22 countries, affecting 500,000 individuals with reported bacterial infections (World Health Organization, 2020).

Indeed, the success and sustainability of the aquaculture industry hinge significantly on effective disease control measures. These measures are crucial for mitigating economic losses attributed to high mortality rates and the costs associated with disease treatment during outbreaks. Minimizing the disease prevalence and severity is crucial for the maintenance of productivity, profitability and effectiveness of aquaculture operations, while safeguarding the well-being of cultured freshwater and marine species (Sciuto et al., 2022). Selective breeding comes in as an alternative solution as it is able to induce cumulative and permanent improvement in the selected trait effectively and profitably (Bachère, 2003; Sciuto et al., 2022).

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 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 to breeding programs in Scotland and Norway to resist IPN genetically (Houston et al., 2008; Moen et al., 2009). Despite its successful application in the aquaculture industry to reduce disease incidence, there was still limited production derived from the selectively bred stocks. However, the proportion was observed to be in the rise rapidly, specifically for high economic value species, including mollusks and crustaceans (Houston et al., 2020). The advancements in methodology and technology have 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 have 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., 2015). Therefore, it is imperative for the continuous monitoring in the changes of pathogen characteristics, including molecular profiles, virulence and antimicrobial resistance (Sciuto et al., 2022).

Pseudomonas koreensis was firstly isolated from Korean agricultural soil (Kwon et al., 2003) and was found as a fish pathogen from golden mahseer (Shahi & Malik, 2014) and empurau (Kho et al., 2023; Lau et al., 2024). Furthermore, genomic and pathogen characterization studies of *P. koreensis* CM-01 had been studied in empurau (Kho et al., 2023; Lau et al., 2024). To date, there are currently no reported disease-resistance markers

associated with *P. koreensis* tolerance for achieving lower mortality rates in *Tor tambroides* through MAS.

4.4.2 T-plex ARMS PCR Assay & Pooled-Sequencing

In this study, the whole genome of two pools of DNA, the resistant group and susceptible group were pooled and sequenced for the identification of disease-resistant SNP 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 (Table 4.2). 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.

Pooled sequencing offers the advantage of predicting allele frequencies of SNP markers with reasonable accuracy, particularly at a 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 & González, 2021). The accuracy of estimating allele frequencies relies on several factors, including the number of pools used, the average sequencing depth, the number of samples combined in each pool, and the potential for variations in the contribution of individual genomes to the final sequencing read (Gautier et al., 2013; Dagnachew et al., 2022). The accuracy of allele frequency estimation via Pearson coefficient had shown up to 0.95 when 2 pools of samples achieved an average of 40X 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 4.2). Nevertheless, the effect

of unequal distribution of individuals into the pool can be decreased by having accurate equimolar pooling of each genomic DNA and balanced number of pooled samples (Gautier et al., 2013; Konczal et al., 2014). An additional difficulty of SNP detection from pooled-sequencing is the difficulty in differentiating rare variants from sequencing errors and this can be solved by applying a minimum threshold allele frequency for the analysis (Guirao-Rico & González, 2021). Pooled sequencing as one of the cost-effective alternatives in SNP marker identification has 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 to 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 (Figure 4.1). 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 are known as outer primers while the primer pairs containing mismatch site (allelic-specific) at their 3' end are the inner primers. Together, the inner primers and the outer primers can simultaneously amplify for both wild-type and the mutant amplicons (Figure 4A). 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, 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 (Figure 4B). 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).

4.4.3 Identification of Disease-Resistant SNPs against *Pseudomonas koreensis*

Single nucleotide polymorphisms (SNPs) are one of the most prevalent genetic variations distributed evenly and widely throughout the genome (Vignal et al., 2002). SNPs can serve as suitable markers for linkage mapping and MAS of desirable traits if the markers are closely linked to the trait (Yue, 2014).

Out of the five selected SNP markers for further validation, all SNP markers were from intron region except for 112388 G/A and 19564 G/C SNP which are positioned in the coding region and intergenic region, respectively (Table 4.4). Studies have emphasized on 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 had been found to enhance mRNA translation efficiency as well (Shaul, 2017).

A study revealed on 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 & Tian, 2016). This possessed a challenge to the scientific community to interpret their involvement in diseases. Researchers had emphasized on the common location of non-coding region near to the regulatory region which further suggesting their possible interference onto the host regulatory elements

(Maurano et al., 2012). Chen & Tian (2016) summarized up to 60.5 % of IGR daSNP-disease phenotype association 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. Up to April 2024, there is still lacking information on the function of DUF5401 superfamily (Wang et al., 2023). Zinc finger family proteins belong to diversified-type of conserved domains to produce novel proteins for lineage-specific processes regulation (Riechmann et al., 2000; Khanna et al., 2009). The family is able to 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 protein that bind to Zn ion to stabilize specialized tertiary structures (Klug & Schwabe, 1995). B-box-type zinc finger domains consist of approximately 40 amino acid residues and can be classified into two main classes: type 1 and type 2 B-box domains. The differences between these two were their consensus sequences and zinc-binding residue arrangement (Short et & Cox, 2006). It can be found within Trim (tripartite motif) proteins which is in good agreement with the high association of the contig to E3 ubiquitin protein ligase Trim35-like. TRIM proteins are characterized by the presence of N-terminal RING finger, one or two B-box domain and a coiled-coil domain (also known as RBCC for Ring, B-box, Coiled-coil) or known as hemopoietic lineage switch or Trim35 (Short & Cox, 2006; Tan et al., 2021). Ubiquitin is characterized as a small modifier molecule specialized for ubiquitination where it labels the abnormal-fold and short-lived proteins for degradation by proteasome (Nandi et al., 2006; Ikeda et al., 2010). In this machinery, E3 ligase played an important role in ubiquitin conjugation machinery due to their substrate specificity which allows them to bind directly to their target proteins (Yang et al., 2009). Trim35 was reported

as a tumor suppressor gene that can inhibit clonogenicity, cell growth and tumorigenicity in HeLa cells and hepatocellular carcinoma, allowing them to play an important role in innate immunity against viral infections (Jia et al., 2011).

Alpha 2C adrenoceptor was characterized within the contig where 112388 G/A SNP marker were chosen from. It belongs to a subcategory of G-protein-coupled receptors (GPCRs) which is responsible for the signal transduction across the eukaryotic cell membrane and its vast range of signals includes light, hormone, olfactory and gustatory stimuli as well as a neurotransmitter (Table 4.4) (Lefkowitz, 2004). GPCRs were found to have around 800 members within humans, comprising the largest family of such receptors. In addition, it is also known as 7TM receptors due to the presence of seven transmembrane helix (7TM). Additionally, the visual pigment rhodopsin was detected within the rod cells, and the ligand-activated β_2 adrenergic receptor (β_2 AR) shares a similar seven-transmembrane helix structure and couples to G proteins (Lefkowitz, 2004). Upon ligand-induced transmembrane signaling, alterations occur on the cytoplasmic surface of the receptor, facilitating the binding of signaling proteins and subsequent catalysis of GTP exchange (Lefkowitz & Shenoy, 2005; Choe et al., 2011). Alpha 2C adrenoceptor belongs to rhodopsin-like GPCR family, which is the largest family of GPCR (class A or rhodopsin-like GPCRs) (Lagerström & Schiöth, 2008). Rhodopsin-like GPCR were found in the human brain as they respond to neurotransmitters. Clinical drugs target mostly on rhodopsin-like GPCR members as well (Choe et al., 2011). Among them, alpha 2C adrenoceptor is reported to be associated with deficit hyperactivity disorder (AHDH) (Cho et al., 2008), behavioral responses including psychiatric disorder and suicide completion (Rivero et al., 2016) as well as colon cancer (Njeim & Eid, 2018).

The contig where 59747 G/A SNP marker is located was characterized as filamin-C-like (FLNC) gene having 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 (Table 4.4) (Stossel et al., 2001). It is widely found across various tissues and cells and not restricted to skeletal and cardiac muscles located at myotendinous junctions, Z-discs, sarcolemma and intercalated discs (Xie et al., 1998). Besides maintaining the mechanical integrity of muscle cells, FLNC was involved in multiple cellular processes including mechanoprotection, intracellular signaling pathways, actin remodeling as well as cell-cell and cell-matrix adhesion (Feng & Walsh, 2004; Fujita et al., 2012). FLNC was found to be associated with myofibrillar myopathy in skeletal and cardiac muscle. Such muscle disorder affecting diverse ethnicities is indeed a mutational hot spot in humans and is due to the truncating mutation in the dimerization domain of FLNC (Fürst et al., 2013). Mutated domains with abnormal structure were less stable which become a seed for abnormal aggregation with other proteins (Kley et al., 2012). Infected individuals would show proximal muscle weakness which further spread to distal and respiratory muscles as the disease progresses (Duff et al., 2011). As the disease commonly arose from the fourth decade of life, it was suggested as an aging-related impairment (Kley et al., 2007). Furthermore, an enlarged heart due to myocardial wall rupture and progressive skeletal muscle degeneration was observed within a medaka (*Oryzias latipes*) mutant with nonsense mutation at FLNC (Fujita et al., 2012).

Fibronectin III and immunoglobulin-like fold (Ig) was characterized as the domain of contactin-5-like (CNTN5) gene found within the contig localized 27857 A/T SNP marker (Table 4.4). CNTN5 as the members of immunoglobulin superfamily of cell adhesion molecules contains six Ig-domains and four fibronectin III domains which is anchored by

glycophosphatidylinositol (GPI)-link to the membrane (Ogawa et al., 1996; Zuko et al., 2013). CNTN5 belongs to contactin family which consisted of altogether six members: contactin 1 to contactin 6. It has been found associated with various neuropsychiatric traits including autism spectrum disorder (ASD) (Fernandez et al., 2004), anorexia nervosa (Nakabayashi et al., 2009), substance abuse (Nikpay et al., 2012) and attention-deficient 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). Further in the year 2015, the role of CNTN5 in cell adhesion is utilized for establishing circuitry and connectivity during early development, as well as for maintaining synaptic connections in adulthood (Kleijer et al., 2015).

Myb/SANT-like DNA binding domain (MSANTD) was found within the contig where 19564 G/C SNP marker is located (Table 4.4). 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 remodeling (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, 19564 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 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 effect 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 27857 A/T SNP showed a single melting temperature (T_m) peak at 76.6 °C for allele C and 75.6 °C for allele T, respectively after subjecting random three samples from each group for Sanger sequencing (Figure 4.3). 112388 G/A and 19564 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 the heterozygous sample with both of its T_m similar to each individual allele (Baris et al., 2013). T-plex real-time PCR assay had been applied in genotyping β -thalassaemia among Iraq patients (Al-Abedy et al., 2019) and introns found on human genes TLR4 and S100A9 (Dhas et al., 2015). The 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).

With the successful identification and validation of 19564 G/C SNP marker (Table 4.5; Figure 4.4) associated with disease resistance towards *Pseudomonas koreensis* in *Tor tambroides* via whole genome pooled sequencing and T-plex ARMS PCR assay, it is hoped that the marker can be used in molecular breeding for the disease-resistant trait of empurau.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

This study reported on the successful assembly of empurau (*Tor tambroides*) genomic data of empurau followed by its characterisation. This is achieved through a hybrid approach combining low-coverage Nanopore long reads and high-coverage Illumina short reads for *de novo* genomic assembly. The resulting draft genome is estimated to be 1.23 Gb in size and exhibits 87.6% completeness based on BUSCO analysis (Actinopterygii_odb10). Predictions yielded 76,419 non-redundant coding sequences, which were further mapped to 304 known KEGG pathways, with signalling transduction being the most prominent pathway. Notably, genes related to growth and immunity were identified, showing significant similarity to previously published genes and holding potential as markers for future molecular breeding efforts in *T. tambroides*. Additionally, the evolutionary relationship of *T. tambroides* with other ray-finned fishes was inferred for the first time using a Maximum-likelihood tree.

Further, this study had also successfully isolated, identified the dominant bacteria of the diseased empurau as *Pseudomonas koreensis* CM-01 strain and further characterised it on the aspect of virulence, growth characteristics, antibiotics susceptibility and disease manifestation. Despite its current infrequent manifestation of disease in aquatic animals, recognizing its pathogenicity remains crucial. Through the characterization of its virulence genes, histopathological observations, antibiotic susceptibility testing, and growth patterns, this chapter offers valuable scientific insights into *P. koreensis*. It serves as a reference for

understanding the pathogen and lays the groundwork for strategies aimed at preventing and treating bacterial diseases caused by *P. koreensis* in fish.

Finally, this study is also reporting on the successful identification and detection of candidate disease-resistant SNP markers towards *Pseudomonas koreensis* in empurau. This was achieved through whole genome pooled sequencing, T-plex ARMS real-time PCR assay and Sanger sequencing on resistant and susceptible pools of fishes. A series of statistical test including F_{ST} pairwise test, Fisher exact test, Cochran-Mantel-Haenszel (CMH) test were with stringent filter criteria were applied for the discovery of candidate SNP markers. With the successful validation of the SNP marker, it is 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 empurau breeding.

5.2 Recommendations

There are still several limitations and future prospects which is required for better improvement and advancement of this study. First, further genomic study for understanding genetic diversity, population structure and evolutionary relationship is recommended via whole genome sequencing of individuals collected from diverse locations, especially the wild-type *T. tambroides*. Secondly, functional genomics involving differential expressed gene under various physiological conditions or in response to pathogen challenge is recommended for future directions as well. Furthermore, marker-assisted breeding programs is recommended for the validated SNP marker to enhance the genetic improvement of empurau in terms of disease resistance toward *P. koreensis*.

REFERENCES

- Aasland, R., Stewart, A. F., & Gibson, T. (1996). The SANT domain: A putative DNA binding domain in the SWI-SNF and ADA complexes, the transcriptional corepressor N-CoR and TFIIIB. *Trends in Biochemical Sciences*, 3(21), 87-88.
- Abdelhamed, H., Ibrahim, I., Baumgartner, W., Lawrence, M. L., & Karsi, A. (2017). Characterization of histopathological and ultrastructural changes in channel catfish experimentally infected with virulent *Aeromonas hydrophila*. *Frontiers in Microbiology*, 8, 1519.
- Abrusán, G., & Krambeck, H. J. (2006). Competition may determine the diversity of transposable elements. *Theoretical Population Biology*, 70(3), 364-375.
- Ahmad, S., Alotaibi, M. A., & Alamri, M. S. (2020). Antibiotic sensitivity pattern of clinical isolates of *Pseudomonas aeruginosa* at a tertiary care hospital in Saudi Arabia: Antibiotic sensitivity pattern of clinical isolates. *Dhaka University Journal of Pharmaceutical Science*, 19(1), 77-82.
- Aizawa, S. I. (2001). Bacterial flagella and type III secretion systems. *FEMS Microbiology Letters*, 202(2), 157-164.
- Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell*, 124(4), 783-801.
- Akmal, Y., Muliari, M., Humairani, R., Zulfahmi, I., Burhanuddin, A. I., Budimawan, B., & Batubara, A. S. (2022). Species authentication of *Tor* spp.(family *Cyprinidae*) in Indonesia based on osteocranium structure and biometric data. *Zoologischer Anzeiger*, 299, 21-30.

- Al-Abedy, N. M., Salman, E. D., & Faraj, S. A. (2019). Frequency of human hemochromatosis HFE gene mutations and serum hepcidin level in iron overload β -thalassaemia Iraqi patients. *Public Health*, 22(10), S275.
- Alday-Sanz, V., Corsin, F., Irde, E., & Bondad-Reantaso, M. G. (2012). Survey on the use of veterinary medicines in aquaculture. *Improving biosecurity through prudent and responsible use of veterinary medicines in aquatic food production*, (pp. 29-44).
- Alexander, M. (1978). Introduction to soil microbiology. *Soil Science*, 125(5), 331.
- Algammal, A. M., Mabrok, M., Sivaramasamy, E., Youssef, F. M., Atwa, M. H., El-Kholy, A. W., Hetta, H. F., & Hozzein, W. N. (2020). Emerging MDR-*Pseudomonas aeruginosa* in fish commonly harbor opr L and tox A virulence genes and bla_{TEM}, bla_{CTX-M}, and tetA antibiotic-resistance genes. *Scientific Reports*, 10(1), 15961.
- Almagor, J., Temkin, E., Benenson, I., Fallach, N., Carmeli, Y., & DRIVE-AB consortium. (2018). The impact of antibiotic use on transmission of resistant bacteria in hospitals: Insights from an agent-based model. *PLOS One*, 13(5), e0197111.
- Al-Tawfiq, J. A. (2007). Occurrence and antimicrobial resistance pattern of inpatient and outpatient isolates of *Pseudomonas aeruginosa* in a Saudi Arabian hospital: 1998-2003. *International Journal of Infectious Diseases*, 11(2), 109-114.
- Altinok, I., Kayis, S., & Capkin, E. (2006). *Pseudomonas putida* infection in rainbow trout. *Aquaculture*, 261(3), 850-855.
- Al-Tobasei, R., Ali, A., Leeds, T. D., 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.

- Alvarez-Pellitero, P. (2008). Fish immunity and parasite infections: From innate immunity to immunoprophylactic prospects. *Veterinary Immunology and Immunopathology*, 126(3-4), 171-198.
- Aly, S. M. (2013). A review of fish diseases in the Egyptian aquaculture sector: Working report.
- Andreozzi, E., & Uhlich, G. A. (2020). PchE regulation of *Escherichia coli* O157: H7 flagella, controlling the transition to host cell attachment. *International Journal of Molecular Sciences*, 21(13), 4592.
- Angelini, N. M., & Seigneur, G. N. (1988). Disease of the fins of *Rhamdia sapo*. Isolation of the etiological agents and experimental infection. *Revista Argentina de Microbiologia*, 20(1), 37-48.
- Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., ... & Brenner, S. (2002). Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, 297(5585), 1301-1310.
- Ardura, A., Linde, A. R., & Garcia-Vazquez, E. (2013). Genetic detection of *Pseudomonas* spp. in commercial amazonian fish. *International Journal of Environmental Research and Public Health*, 10(9), 3954-3966.
- Arkoosh, M. R., & Kaattari, S. L. (1991). Development of immunological memory in rainbow trout (*Oncorhynchus mykiss*). I. An immunochemical and cellular analysis of the B cell response. *Developmental & Comparative Immunology*, 15(4), 279-293.
- Austin, B., & Austin, D. A. (2016). Pseudomonads. In: *Bacterial Fish Pathogens*. (pp 475 498). Springer International Publishing.

- Austin, B., Austin, D. A., & Munn, C. B. (2007). *Bacterial fish pathogens: Disease of farmed and wild fish* (Vol. 26, p. 552). Dordrecht, The Netherlands: Springer.
- Austin, C. M., Tan, M. H., Harrison, K. A., Lee, Y. P., Croft, L. J., Sunnucks, P., ... & Gan, H. M. (2017). *De novo* genome assembly and annotation of Australia's largest freshwater fish, the Murray cod (*Maccullochella peelii*), from Illumina and Nanopore sequencing read. *GigaScience*, 6(8), gix063.
- Babalola, O. O., & Ayangbenro, A. S. (2019). Draft genome sequence of *Pseudomonas koreensis* strain AB36, isolated from gold mining soil. *Microbiology Resource Announcements*, 8(20), 10-1128.
- Bachère, E. (2003). Anti-infectious immune effectors in marine invertebrates: Potential tools for disease control in larviculture. *Aquaculture*, 227(1-4), 427-438.
- Bai, J. S., Li, Y. W., Deng, Y., Huang, Y. Q., He, S. H., Dai, J., Zhao, S. Z., Dan, X. M., & Luo, X. C. (2017). Molecular identification and expression analysis of TLR5M and TLR5S from orange spotted grouper (*Epinephelus coioides*). *Fish & Shellfish Immunology*, 63, 97-102.
- Bajpai, P. K., Harel, A., Shafir, S., & Barazani, O. (2022). Whole genome sequencing reveals footprints of adaptive genetic variation in populations of *Eruca sativa*. *Frontiers in Ecology and Evolution*, 10, 938981.
- Bakon, S. K., Mohamad, Z. A., Jamilan, M. A., Hashim, H., Kuman, M. Y., Shaharudin, R., Ahmad, N., & Muhamad, N. A. (2023). Prevalence of antibiotic-resistant pathogenic bacteria and level of antibiotic residues in hospital effluents in Selangor, Malaysia: Protocol for a cross-sectional study. *JMIR Research Protocols*, 12(1), e39022.

- Ballesterio, J. G. D. A., Garcia, J. M., Bollela, V. R., Ruffino-Netto, A., Dalcolmo, M. M. P., Moncaio, A. C. S., Miguel, N. S., Rigolin, I. Z., & Palha, P. F. (2020). Management of multidrug-resistant tuberculosis: Main recommendations of the Brazilian guidelines. *Jornal Brasileiro de Pneumologia*, 46, e20190290.
- Baltrušis, P., Doyle, S. R., Halvarsson, P., & Höglund, J. (2022). Genome-wide analysis of the response to ivermectin treatment by a Swedish field population of *Haemonchus contortus*. *International Journal for Parasitology: Drugs and Drug Resistance*, 18, 12-19.
- Bankers, L., Fields, P., McElroy, K. E., Boore, J. L., Logsdon Jr, J. M., & Neiman, M. (2017). Genomic evidence for population-specific responses to co-evolving parasites in a New Zealand freshwater snail. *Molecular Ecology*, 26(14), 3663-3675.
- Barasch, N., Gong, X., Kwei, K. A., Varma, S., Biscocho, J., Qu, K., Xiao, N., Lipsick, J. S., Pelham, R. J., West, R. B., & 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.
- Bardoel, B. W., van Kessel, K. P., van Strijp, J. A., & Milder, F. J. (2012). Inhibition of *Pseudomonas aeruginosa* virulence: Characterization of the AprA–AprI interface and species selectivity. *Journal of Molecular Biology*, 415(3), 573-583.
- Baris, I., Etlik, O., Koksall, V., Ocak, Z., & Baris, S. T. (2013). SYBR green dye-based probe free SNP genotyping: Introduction of T-Plex real-time PCR assay. *Analytical Biochemistry*, 441(2), 225-231.

- Barnett, C. W., & Pankhurst, N. W. (1998). The effects of common laboratory and husbandry practices on the stress response of greenback flounder *Rhombosolea tapirine* (Günther, 1862). *Aquaculture*, 162(3-4), 313-329.
- Basso, P., Wallet, P., Elsen, S., Soleilhac, E., Henry, T., Faudry, E., & Attrée, I. (2017). Multiple *Pseudomonas* species secrete exolysin-like toxins and provoke Caspase-1-dependent macrophage death. *Environmental Microbiology*, 19(10), 4045-4064.
- Basu, M., Paichha, M., Swain, B., Lenka, S. S., Singh, S., Chakrabarti, R., & Samanta, M. (2015). Modulation of TLR2, TLR4, TLR5, NOD1 and NOD2 receptor gene expressions and their downstream signaling molecules following thermal stress in the Indian major carp catla (*Catla catla*). *3 Biotech*, 5(6), 1021-1030.
- Bayne, C. J., & Gerwick, L. (2001). The acute phase response and innate immunity of fish. *Developmental & Comparative Immunology*, 25(8-9), 725-743.
- Benie, C. K. D., Dadié, A., Guessennd, N., N'gbesso-Kouadio, N. A., Kouame, N. Z. D., N'golo, D. C., Aka, S., Dako, E., Dje, K. M., & Dosso, M. (2017). Characterization of virulence potential of *Pseudomonas aeruginosa* isolated from bovine meat, fresh fish, and smoked fish. *European Journal of Microbiology and Immunology*, 7(1), 55-64.
- Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M., Beacham, T. D., ... & Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution*, 32(9), 665-680.
- Bhardwaj, S., Mehra, P., Dhanjal, D. S., Sharma, P., Sharma, V., Singh, R., Nepovimova, E., Chopra, C., & Kua, K. (2022). Antibiotics and antibiotic resistance-flipsides of the same coin. *Current Pharmaceutical Design*, 28(28), 2312-2329.

- Blazer, V. S. (1991). Piscine macrophage function and nutritional influences: A review. *Journal of Aquatic Animal Health*, 3(2), 77-86.
- Bly, J. E., & Clem, L. W. (1992). Temperature and teleost immune functions. *Fish & Shellfish Immunology*, 2(3), 159-171.
- Boitard, S., Schlötterer, C., Nolte, V., Pandey, R. V., & Futschik, A. (2012). Detecting selective sweeps from pooled next-generation sequencing samples. *Molecular Biology and Evolution*, 29(9), 2177-2186.
- Bondad-Reantaso, M. G., MacKinnon, B., Karunasagar, I., Fridman, S., Alday-Sanz, V., Brun, E., Li, A., Surachetpong, W., Karunasagar, I., Hao, B., Dall'Occo, A., Urbani, R., & Caputo, A. (2023). Review of alternatives to antibiotic use in aquaculture. *Reviews in Aquaculture*, 15(4), 1421-1451.
- Borodulina, O. R., Golubchikova, J. S., Ustyantsev, I. G., & Kramerov, D. A. (2016). Polyadenylation of RNA transcribed from mammalian SINEs by RNA polymerase III: Complex requirements for nucleotide sequences. *Biochimica et Biophysica Acta (BBA)- Gene Regulatory Mechanisms*, 1859(2), 355-365.
- Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., ... & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology*, 19, 1-12.
- Bowden, T. J., Butler, R., Bricknell, I. R., & Ellis, A. E. (1997). Serum trypsin-inhibitory activity in five species of farmed fish. *Fish & Shellfish Immunology*, 7(6), 377-385.
- Bowden, T. J., Cook, P., & Rombout, J. H. W. M. (2005). Development and function of the thymus in teleosts. *Fish & Shellfish Immunology*, 19(5), 413-427.

- Boyer, L. A., Latek, R. R., & Peterson, C. L. (2004). The SANT domain: a unique histone-tail binding module?. *Nature Reviews Molecular Cell Biology*, 5(2), 158-163.
- Breuil, G., Vassiloglou, B., Pepin, J. F., & Romestand, B. (1997). Ontogeny of IgM-bearing cells and changes in the immunoglobulin M-like protein level (IgM) during larval stages in sea bass (*Dicentrarchus labrax*). *Fish & Shellfish Immunology*, 7(1), 29-43.
- Brosius, J. (1999). RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene*, 238(1), 115-134.
- Bruna, T., Hoff, K.J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database. *NAR Genomics and Bioinformatics*, 3(1).
- Budd, A. M., Banh, Q. Q., Domingos, J. A., & Jerry, D. R. (2015). Sex control in fish: Approaches, challenges and opportunities for aquaculture. *Journal of Marine Science and Engineering*, 3(2), 329-355.
- Buonocore, F., Randelli, E., Bird, S., Secombes, C. J., Costantini, S., Facchiano, A., Mazzini, M., & Scapigliati, G. (2006). The CD8 α from sea bass (*Dicentrarchus labrax* L.): Cloning, expression and 3D modelling. *Fish & Shellfish Immunology*, 20(4), 637-646.
- Busse, H. J., Denner, E. B., & Lubitz, W. (1996). Classification and identification of bacteria: Current approaches to an old problem. Overview of methods used in bacterial systematics. *Journal of Biotechnology*, 47(1), 3-38.
- Cabello, F. C. (2006). Heavy use of prophylactic antibiotics in aquaculture: A growing problem for human and animal health and for the environment. *Environmental Microbiology*, 8(7), 1137-1144.

- Cantalapiedra, C. P., 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. *Molecular Biology and Evolution*, 38(12), 5825-5829.
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972-1973.
- Carr, I. M., Robinson, J. I., Dimitriou, R., Markham, A. F., Morgan, A. W., & Bonthron, D. T. (2009). Inferring relative proportions of DNA variants from sequencing electropherograms. *Bioinformatics*, 25(24), 3244-3250.
- Carroll, M. C. (2004). The complement system in B cell regulation. *Molecular Immunology*, 41(2-3), 141-146.
- Carroll, M. C., & Janeway Jr, C. A. (1999). Innate immunity: Editorial overview. *Current Opinion in Immunology*, 11(1), 11-12.
- Castric, P. A. (1983). Hydrogen cyanide production by *Pseudomonas aeruginosa* at reduced oxygen levels. *Canadian Journal of Microbiology*, 29(10), 1344-1349.
- Castric, P. A., Ebert, R. F., & Castric, K. F. (1979). The relationship between growth phase and cyanogenesis in *Pseudomonas aeruginosa*. *Current Microbiology*, 2, 287-292.
- Chandhini, S., & Rejish Kumar, V. J. (2019). Transcriptomics in aquaculture: Current status and applications. *Reviews in Aquaculture*, 11(4), 1379-1397.
- Chatterjee, A., Cui, Y., Yang, H., Collmer, A., Alfano, J. R., & Chatterjee, A. K. (2003). GacA, the response regulator of a two-component system, acts as a master regulator in *Pseudomonas syringae* pv. tomato DC3000 by controlling regulatory RNA,

- transcriptional activators, and alternate sigma factors. *Molecular Plant-Microbe Interactions*, 16(12), 1106-1117.
- Chen, F., Sun, J., Han, Z., Yang, X., Xian, J. A., Lv, A., Hu, X., & Shi, H. (2019a). Isolation, identification and characteristics of *Aeromonas veronii* from diseased crucian carp (*Carassius auratus gibelio*). *Frontiers in Microbiology*, 10, 2742.
- Chen, J., & Tian, W. (2016). Explaining the disease phenotype of intergenic SNP through predicted long range regulation. *Nucleic Acids Research*, 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.
- Chen, Z., Omori, Y., Koren, S., Shirokiya, T., Kuroda, T., Miyamoto, A., Wada, H., Fujiyama, A., Toyoda, A., Zhang, S., Wolfsberg, T. G., Kawakami, K., Phillippy, A. D., Program, N. C. S., Mullikin, J. C., & Burgess, S. M. (2019b). *De novo* assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole genome duplication. *Science Advances*, 5(6), eaav0547.
- 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, M., Kim, S. A., Cho, D. Y., Yoo, H. J., Chung, U. S., Son, J. W., & 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.

- Choe, H. W., Park, J. H., Kim, Y. J., & Ernst, O. P. (2011). Transmembrane signaling by GPCRs: insight from rhodopsin and opsin structures. *Neuropharmacology*, 60(1), 52-57.
- Chokshi, A., Sifri, Z., Cennimo, D., & Horng, H. (2019). Global contributors to antibiotic resistance. *Journal of Global Infectious Diseases*, 11(1), 36-42.
- Chong, P. P., Adnan, A. M., & Wan, K. L. (2011). Characterisation of simple sequence repeats in the Asian seabass, *Lates calcarifer* by random sequencing. *Sains Malaysiana*, 40.
- Christen, G. L., & Marshall, R. T. (1984). Selected properties of lipase and protease of *Pseudomonas fluorescens* 27 produced in four media. *Journal of Dairy Science*, 67(8), 1680-1687.
- Chung, H. H. (2018). Real-time polymerase chain reaction (RT-PCR) for the authentication of raw meats.
- Churchill, M. E., & Chen, L. (2011). Structural basis of acyl-homoserine lactone-dependent signaling. *Chemical Reviews*, 111(1), 68-85.
- Collins, A., & Ke, X. (2012). Primer1: Primer design web service for tetra-primer ARMS-PCR. *The Open Bioinformatics Journal*, 6, 55-58.
- Crovadore, J., Cochard, B., Chablais, R., Haenzi, M., Raffini, F., & Lefort, F. (2020). Draft genome sequences of *Pseudomonas koreensis* strain UASWS1668, *Bacillus megaterium* strain UASWS1667, and *Paenibacillus* sp. strain UASWS1643, considered potential plant growth-promoting rhizobacteria. *Microbiology Resource Announcements*, 9(33), 10-1128.

- Cuesta, A., Vargas-Chacoff, L., García-López, A., Arjona, F. J., Martínez-Rodríguez, G., Meseguer, J., Mancera, J. M., & Esteban, M. A. (2007). Effect of sex-steroid hormones, testosterone and estradiol, on humoral immune parameters of gilthead seabream. *Fish & Shellfish Immunology*, 23(3), 693-700.
- Dagnachew, B., Aslam, M. L., 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. *Frontiers in Genetics*, 13, 896774.
- Dalmo, R. A. (2005). Ontogeny of the fish immune system. *Fish and Shellfish Immunology*, 5(19), 395-396.
- Dam, C. T. M., Ventura, T., Booth, M., Pirozzi, I., Salini, M., Smullen, R., & Elizur, A. (2020). Intestinal transcriptome analysis highlights key differentially expressed genes involved in nutrient metabolism and digestion in yellowtail kingfish (*Seriola lalandi*) fed terrestrial animal and plant proteins. *Genes*, 11(6), 621.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, 10(2), giab008.
- Danilova, N., Bussmann, J., Jekosch, K., & Steiner, L. A. (2005). The immunoglobulin heavy-chain locus in zebrafish: Identification and expression of a previously unknown isotype, immunoglobulin Z. *Nature Immunology*, 6(3), 295-302.
- Danzmann, R. G., Kocmarek, A. L., Norman, J. D., Rexroad, C. E., & Palti, Y. (2016). Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. *BMC Genomics*, 17(1), 1-18.

- Demers, N. E., & Bayne, C. J. (1997). The immediate effects of stress on hormones and plasma lysozyme in rainbow trout. *Developmental & Comparative Immunology*, 21(4), 363-373.
- Denning, D. W., Perlin, D. S., Muldoon, E. G., Colombo, A. L., Chakrabarti, A., Richardson, M. D., & Sorrell, T. C. (2017). Delivering on antimicrobial resistance agenda not possible without improving fungal diagnostic capabilities. *Emerging Infectious Diseases*, 23(2), 177.
- Department of Fisheries Malaysia (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, D. B. B., Ashmi, A. H., Bhat, B. V., Parija, S. C., & 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 Journal of Medical Research*, 142(5), 555-562.
- Di Genova, A., Buena-Atienza, E., Ossowski, S., & Sagot, M. F. (2021). Efficient hybrid *de novo* assembly of human genomes with WENGAN. *Nature Biotechnology*, 39(4), 422-430.
- Dixon, B., van Erp, S. H., Rodrigues, P. N., Egberts, E., & Stet, R. M. (1995). Fish major histocompatibility complex genes: An expansion. *Developmental & Comparative Immunology*, 19(2), 109-133.

- Dodds, D. R. (2017). Antibiotic resistance: A current epilogue. *Biochemical Pharmacology*, 134, 139-146.
- Donsakul, T., & Magtoon, W. (2008). Karyotypes of seven cyprinid fishes: *Systomus binotatus*, *Puntius brevis*, *Poropuntius laoensis*, *Labiobarbus siamensis*, *Catlocarpio siamensis*, *Tor tambroides* and *Probarbus jullieni* from Thailand. *Srinakharinwirot Science Journal*, 24, 80-92.
- Du Pasquier, L. (1982). Antibody diversity in lower vertebrates—why is it so restricted?. *Nature*, 296(5855), 311-313.
- Du Pasquier, L. (2001). The immune system of invertebrates and vertebrates. *Comparative Biochemistry and Physiology. Part B, Biochemistry & Molecular Biology*, 129(1), 1-15.
- Du Pasquier, L. (2004). Innate immunity in early chordates and the appearance of adaptive immunity. *Comptes rendus biologies*, 327(6), 591-601.
- Duff, R. M., Tay, V., Hackman, P., Ravenscroft, G., McLean, C., Kennedy, P., Steinbach, A., Schoffler, W., van der Ven, P. F. M., Furst, D. O., Song J., DjinoVIC-Carugo, K., Penttila, S., Raheem, O., Reardon, K., Malandrini, A., Gambelli, S., Villanova, M., Nowak, K. J., Williams, D. R., Landers, J. E., Brown, R. H., Udd, B., & Laing, N. G. (2011). Mutations in the N-terminal actin-binding domain of filamin C cause a distal myopathy. *The American Journal of Human Genetics*, 88(6), 729-740.
- Duman, M., Mulet, M., Altun, S., Satıcıoğlu, I. B., Özdemir, B., Ajmi, N., Lalucat, J., & García Valdés, E. (2021). The diversity of *Pseudomonas* species isolated from fish farms in Turkey. *Aquaculture*, 535, 736369.

- Durand, G. A., Raoult, D., & Dubourg, G. (2019). Antibiotic discovery: History, methods and perspectives. *International Journal of Antimicrobial Agents*, 53(4), 371-382.
- Eckert, K. A., Yan, G., & Hile, S. E. (2002). Mutation rate and specificity analysis of tetranucleotide microsatellite DNA alleles in somatic human cells. *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center*, 34(3), 140-150.
- Economou, V., & Gousia, P. (2015). Agriculture and food animals as a source of antimicrobial resistant bacteria. *Infection and Drug Resistance*, 49-61.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792-1797.
- Eickbush, T. H., & Jamburuthugoda, V. K. (2008). The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Research*, 134(1-2), 221-234.
- Eissa, N. M. E., El-Ghiet, E. A., Shaheen, A. A., & Abbass, A. (2010). Characterization of *Pseudomonas* species isolated from tilapia “*Oreochromis niloticus*” in Qaroun and Wadi-El-Rayan lakes, Egypt. *Global Veterinaria*, 5(2), 116-121.
- Ellingsen, T., Strand, C., Monsen, E., Bøgwald, J., & Dalmo, R. A. (2005). The ontogeny of complement component C3 in the spotted wolffish (*Anarhichas minor* Olafsen). *Fish & Shellfish Immunology*, 18(5), 351-358.
- Ellis, A. E. (1999). Immunity to bacteria in fish. *Fish & Shellfish Immunology*, 9(4), 291-308.
- Ellis, A. E. (2001). Innate host defense mechanisms of fish against viruses and bacteria. *Developmental & Comparative Immunology*, 25(8-9), 827-839.

- Elsen, S., Huber, P., Bouilliot, S., Couté, Y., Fournier, P., Dubois, Y., Timisit, J., Maurin, M., & Attrée, I. (2014). A type III secretion negative clinical strain of *Pseudomonas aeruginosa* employs a two partner secreted exolysin to induce hemorrhagic pneumonia. *Cell Host & Microbe*, 15(2), 164-176.
- Espelid, S., Løkken, G. B., Steiro, K., & Bøgwald, J. (1996). Effects of cortisol and stress on the immune system in Atlantic Salmon (*Salmo salar* L.). *Fish & Shellfish Immunology*, 6(2), 95-110.
- 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.
- European Centre for Disease Prevention and Control. (2020). *Antimicrobial resistance in the EU/EEA (EARS-Net)—annual epidemiological report 2019*. ECDC, Stockholm.
- Evans, D. L., & Jaso-Friedmann, L. (1992). Nonspecific cytotoxic cells as effectors of immunity in fish. *Annual Review of Fish Diseases*, 2, 109-121.
- Fadhil, L., Al-Marzoqi, A. H., Al Taei, Z. M., & Shalan, A. A. (2016). Molecular and phenotypic study of virulence genes in a pathogenic strain of *Pseudomonas aeruginosa* isolated from various clinical origins by PCR: Profiles of genes and toxins. *Research Journal of Pharmaceutical Biological and Chemical Sciences*, 7(1), 590-598.
- Fan, G., Song, Y., Yang, L., Huang, X., Zhang, S., Zhang, M., ... & He, S. (2020). Initial data release and announcement of the 10,000 Fish Genomes Project (Fish10K). *GigaScience*, 9(8), giaa080.

- Fang, H. M., Ge, R., & Sin, Y. M. (2004). Cloning, characterisation and expression of *Aeromonas hydrophila* major adhesin. *Fish & Shellfish Immunology*, 16(5), 645-658.
- 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>
- FAO. (2018). The state of world fisheries and aquaculture 2018 - meeting the sustainable development goals. (Rome).
- FAO. (2020a). Fishery and Aquaculture Country Profiles. Malaysia. Food and Agriculture Organization of the United Nations. Retrieved from <http://www.fao.org/fishery/facp/MYS/en>.
- FAO. (2020b). Fisheries statistical collections: Global aquaculture production. food and agriculture organization of the United Nations. Retrieved from <http://www.fao.org/fishery/statistics/global-aquaculture-production/en>.
- FAO. (2021). *Thematic background study – genome-based biotechnologies in aquaculture*. Rome. <https://doi.org/10.4060/cb7955en>
- Fast, M. D., Hosoya, S., Johnson, S. C., & Afonso, L. O. (2008). Cortisol response and immune-related effects of Atlantic salmon (*Salmo salar* Linnaeus) subjected to short and long-term stress. *Fish & Shellfish Immunology*, 24(2), 194-204.
- Fazeli, N., & Momtaz, H. (2014). Virulence gene profiles of multidrug-resistant *Pseudomonas aeruginosa* isolated from Iranian hospital infections. *Iranian Red Crescent Medical Journal*, 16(10).

- Feng, K., Li, R., Chen, Y., Zhao, B., & Yin, T. (2015). Sequencing and analysis of the *Pseudomonas fluorescens* GcM5-1A genome: A pathogen living in the surface coat of *Bursaphelenchus xylophilus*. *PLoS One*, 10(10), e0141515.
- Feng, Y., & Walsh, C. A. (2004). The many faces of filamin: A versatile molecular scaffold for cell motility and signalling. *Nature Cell Biology*, 6(11), 1034-1038.
- Fernandez, T., Morgan, T., Davis, N., Klin, A., Morris, A., Farhi, A., & Lifton, R. P. (2004). Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. *The American Journal of Human Genetics*, 74(6), 1286-1293.
- Feschotte, C., & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics*, 41(1), 331-368.
- Figueras, A., Robledo, D., Corvelo, A., Hermida, M., Pereiro, P., Rubiolo, J. A., Gomez Garrido, J., Carrete, L., Bello, X., Gut, M., Gut, I. G., Marcet-Houben, M., Forn Cuni, G., Galan, B., Garcia, J. L., Abal-Fabeiro, J. L., Pardo, B. G., Taboada, X., Fernandez, C., Vlasova, A., Hermoso-Pulido, A., Fugo, R., Alvarez-Dios, J. A., Gomez-Tato, A., Vinas, A., Maside, X., Gabaldon, T., Novoa, B., Bouza, C., Alioto, T., & Martínez, P. (2016). Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): A fish adapted to demersal life. *DNA Research*, 23(3), 181-192.
- Fink, I. R., Pietretti, D., Voogdt, C. G., Westphal, A. H., Savelkoul, H. F., Forlenza, M., & Wiegertjes, G. F. (2016). Molecular and functional characterization of Toll-like receptor (Tlr) 1 and Tlr2 in common carp (*Cyprinus carpio*). *Fish & Shellfish Immunology*, 56, 70-83.

- Firdaus-Nawi, M., and Zamri-Saad, M. (2016). Major components of fish immunity: A review. *Pertanika Journal of Tropical Agricultural Science*, 39, 393–420.
- Fischer, M. C., Rellstab, C., Tedder, A., Zoller, S., Gugerli, F., Shimizu, K. K., Holderegger, R., & 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.
- Fischer, U., Utke, K., Somamoto, T., Köllner, B., Ototake, M., & Nakanishi, T. (2006). Cytotoxic activities of fish leucocytes. *Fish & Shellfish Immunology*, 20(2), 209-226.
- Fluit, A. C., Rogers, M. R., Díez-Aguilar, M., Cantón, R., Benaissa-Trouw, B. J., Bayjanov, J. R., & Ekkelenkamp, M. B. (2020). Draft genome sequence of the strain 16-537536, isolated from a patient with bronchiectasis and its relationship to the *Pseudomonas koreensis* group of the *Pseudomonas fluorescens* complex. *BMC Research Notes*, 13(1), 1-5.
- Franch, R., Cardazzo, B., Antonello, J., Castagnaro, M., Patarnello, T., & Bargelloni, L. (2006). Full-length sequence and expression analysis of Toll-like receptor 9 in the gilthead seabream (*Sparus aurata* L.). *Gene*, 378, 42-51.
- Fu, G. H., Liu, F., Xia, J. H., & Yue, G. H. (2014). The LBP gene and its association with resistance to *Aeromonas hydrophila* in tilapia. *International Journal of Molecular Sciences*, 15(12), 22028-22041.
- Fujita, M., Mitsunashi, H., Isogai, S., Nakata, T., Kawakami, A., Nonaka, I., Noguchi, S., Hayashi, Y. K., Nishino, 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.

- Fürst, D. O., Goldfarb, L. G., Kley, R. A., Vorgerd, M., Olivé, M., & van der Ven, P. F. (2013). Filamin C-related myopathies: Pathology and mechanisms. *Acta Neuropathologica*, 125, 33-46.
- Gabriel, L., Hoff, K. J., Bruna, T., Borodovsky, M., & Stanke, M. (2021). TSEBRA: Transcript selector for BRAKER. *BMC Bioinformatics*, 22(1), 1-12.
- Gajdács, M., & Albericio, F. (2019). Antibiotic resistance: From the bench to patients. *Antibiotics*, 8(3), 129.
- Gao, B., Wang, Y., Diaby, M., Zong, W., Shen, D., Wang, S., ... & Song, C. (2020). Evolution of pogo, a separate superfamily of IS630-Tc1-mariner transposons, revealing recurrent domestication events in vertebrates. *Mobile DNA*, 11, 1-15.
- Garrido-Sanz, D., Meier-Kolthoff, J. P., Göker, M., Martín, M., Rivilla, R., & Redondo Nieto, M. (2016). Genomic and genetic diversity within the *Pseudomonas fluorescens* complex. *PloS one*, 11(2), e0150183.
- Gautier, M., Foucaud, J., Gharbi, K., Cézard, T., Galan, M., Loiseau, A., Thomson, M., Pudlo, P., Kerdelhue, C., & 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. *Marine biotechnology*, 19, 570-578.
- Gess, R. W., Coates, M. I., & Rubidge, B. S. (2006). A lamprey from the Devonian period of South Africa. *Nature*, 443(7114), 981-984.

- Girlich, D., Poirel, L., & Nordmann, P. (2010). Novel Ambler class A carbapenem hydrolyzing β -lactamase from a *Pseudomonas fluorescens* isolate from the Seine River, Paris, France. *Antimicrobial Agents and Chemotherapy*, 54(1), 328-332.
- Gornati, R., Papis, E., Rimoldi, S., Terova, G., Saroglia, M., & Bernardini, G. (2004). Rearing density influences the expression of stress-related genes in sea bass (*Dicentrarchus labrax*, L.). *Gene*, 341, 111-118.
- Govarthanan, M., Selvankumar, T., Mythili, R., Srinivasan, P., Ameen, F., AlYahya, S. A., & Kamala-Kannan, S. (2019). Biogreen remediation of chromium-contaminated soil using *Pseudomonas* sp.(RPT) and neem (*Azadirachta indica*) oil cake. *International Journal of Environmental Science and Technology*, 16, 4595-4600.
- Grinde, B., Lie, Ø., Poppe, T., & Salte, R. (1988). Species and individual variation in lysozyme activity in fish of interest in aquaculture. *Aquaculture*, 68(4), 299-304.
- Guan, W. Z., & Qiu, G. F. (2020). Transcriptome analysis of the growth performance of hybrid mandarin fish after food conversion. *PLOS One*, 15(10), e0240308.
- Guirao-Rico, S., & González, J. (2021). Benchmarking the performance of Pool-seq SNP callers using simulated and real sequencing data. *Molecular Ecology Resources*, 21(4), 1216-1229.
- Gutenberger, S. K., Duimstra, J. R., Rohovec, J. S., & Fryer, J. L. (1997). Intracellular survival of *Renibacterium salmoninarum* in trout mononuclear phagocytes. *Diseases of Aquatic Organisms*, 28(2), 93-106.
- Hamid, N. H., Hassan, M. D., Sabri, M. M., Hasliza, A. H., Hamdan, R. H., Afifah, M. N., Manfa, S. R., Nadia, A. B., & Fuad, M. M. (2017). Studies on pathogenicity effect of *Aeromonas hydrophila* infection in juvenile red hybrid tilapia *Oreochromis*

- sp. In *International Seminar on Livestock Production and Veterinary Technology* (pp. 532-539).
- Han, B., Liang, W., Liu, L., Li, Y., & Sun, D. (2018). Genetic association of the ACACB gene with milk yield and composition traits in dairy cattle. *Animal Genetics*, 49(3), 169-177.
- Han, Z., Sun, J., Lv, A., Sung, Y., Shi, H., Hu, X., & Xing, K. (2017). Isolation, identification and characterization of *Shewanella algae* from reared tongue sole, *Cynoglossus semilaevis* Günther. *Aquaculture*, 468, 356-362.
- Hanif, A., Bakopoulos, V., Leonardos, I., & Dimitriadis, G. J. (2005). The effect of sea bream (*Sparus aurata*) broodstock and larval vaccination on the susceptibility by *Photobacterium damsela* subsp. *piscicida* and on the humoral immune parameters. *Fish & Shellfish Immunology*, 19(4), 345-361.
- Hansen, J. D., Landis, E. D., & Phillips, R. B. (2005). Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish. *Proceedings of the National Academy of Sciences*, 102(19), 6919-6924.
- Harding, C. V., & Neefjes, J. (2005). Antigen processing and recognition. *Current Opinion in Immunology*, 1(17), 55-57.
- Hauser, A. R. (2009). The type III secretion system of *Pseudomonas aeruginosa*: Infection by injection. *Nature Reviews Microbiology*, 7(9), 654-665.
- Hayward, C. J., Bott, N. J., & Nowak, B. F. (2009). Seasonal epizootics of sea lice, *Caligus* spp., on southern bluefin tuna, *Thunnus maccoyii* (Castelnau), in a long-term farming trial. *Journal of Fish Diseases*, 32(1), 101-106.

- He, L. B., Wang, H., Luo, L. F., Jiang, S. H., Liu, L. Y., Li, Y. M., Huang, R., Liao, L. J., Zhu, Z. Y., & Wang, Y. P. (2016). Characterization, expression analysis and localization pattern of toll-like receptor 1 (tlr1) and toll-like receptor 2 (tlr2) genes in grass carp *Ctenopharyngodon idella*. *Journal of Fish Biology*, 89(2), 1434-1440.
- He, Y., Xu, T., Fossheim, L. E., & Zhang, X. H. (2012). FliC, a flagellin protein, is essential for the growth and virulence of fish pathogen *Edwardsiella tarda*. *PLOS ONE*, 7(9).
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1-8.
- Herbomel, P., Thisse, B., & Thisse, C. (1999). Ontogeny and behaviour of early macrophages in the zebrafish embryo. *Development*, 126(17), 3735-3745.
- Herbomel, P., Thisse, B., & Thisse, C. (2001). Zebrafish early macrophages colonize cephalic mesenchyme and developing brain, retina, and epidermis through a M-CSF receptor-dependent invasive process. *Developmental biology*, 238(2), 274-288.
- Høiby, N., Bjarnsholt, T., Givskov, M., Molin, S., & Ciofu, O. (2010). Antibiotic resistance of bacterial biofilms. *International Journal of Antimicrobial Agents*, 35(4), 322-332.
- Holland, J. W., Pottinger, T. G., & Secombes, C. J. (2002). Recombinant interleukin-1beta activates the hypothalamic-pituitary-interrenal axis in rainbow trout, *Oncorhynchus mykiss*. *Journal of Endocrinology*, 175(1), 261-267.
- Holland, M. C. H., & Lambris, J. D. (2002). The complement system in teleosts. *Fish & Shellfish Immunology*, 12(5), 399-420.

- Holmes, A. H., Moore, L. S., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P. J., & Piddock, L. J. (2016). Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet*, 387(10014), 176-187.
- Hong, J. (1998). The distribution trends in simple repetitive stretches of DNA. *Chinese Journal of Biochemistry and Molecular Biology*, 14(1), 65-70.
- Hosoe, A., Suenaga, T., Sugi, T., Iizumi, T., Nagai, N., & Terada, A. (2020). Complete genome sequence of *Pseudomonas putida* strain TS312, harboring an HdtS-type N-acyl-homoserine lactone synthase, isolated from a paper mill. *Microbiology Resource Announcements*, 9(13), 10-1128.
- Houston, R. D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., Selly, S. C., Martin, S. A. M., Stevens, J. R., Santos, E. M., Davie, A., & Robledo, D. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews Genetics*, 21(7), 389-409.
- Houston, R. D., Gheyas, A., Hamilton, A., Guy, D. R., Tinch, A. E., Taggart, J. B., McAndrew, B. J., Haley, C. S., & 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.
- Hu, G., Gu, W., Sun, P., Bai, Q., & Wang, B. (2016). Transcriptome analyses reveal lipid metabolic process in liver related to the difference of carcass fat content in rainbow trout (*Oncorhynchus mykiss*). *International Journal of Genomics*, 2016.
- Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, 36(7), 2253-2255.

- Huang, M., Chen, Y., Han, D., Lei, Z., & Chu, X. (2019). Role of the zinc finger and SCAN domain-containing transcription factors in cancer. *American Journal of Cancer Research*, 9(5), 816.
- Huang, R., Dong, F., Jang, S., Liao, L., Zhu, Z., & Wang, Y. (2012). Isolation and analysis of a novel grass carp toll-like receptor 4 (tlr4) gene cluster involved in the response to grass carp reovirus. *Developmental & Comparative Immunology*, 38(2), 383-388.
- Huttenhuis, H. B., Huising, M. O., van der Meulen, T., van Oosterhoud, C. N., Sánchez, N. A., Taverne-Thiele, A. J., Stroband, H. W., & Rombout, J. H. (2005). Rag expression identifies B and T cell lymphopoietic tissues during the development of common carp (*Cyprinus carpio*). *Developmental & Comparative Immunology*, 29(12), 1033-1047.
- Hwang, S. D., Kondo, H., Hirono, I., & Aoki, T. (2011). Molecular cloning and characterization of Toll-like receptor 14 in Japanese flounder, *Paralichthys olivaceus*. *Fish & Shellfish Immunology*, 30(1), 425-429.
- Ikeda, F., Crosetto, N., & Dikic, I. (2010). What determines the specificity and outcomes of ubiquitin signaling?. *Cell*, 143(5), 677-681.
- Ingram, B., Sungan, S., Gooley, G., Sim, S. Y., Tinggi, D., & De Silva, S. S. (2005). Induced spawning, larval development and rearing of two indigenous Malaysian mahseer, *Tor tambroides* and *Tor douronensis*. *Aquaculture Research*, 36(10), 983-995.
- Irshath, A. A., Rajan, A. P., Vimal, S., Prabhakaran, V. S., & Ganesan, R. (2023). Bacterial pathogenesis in various fish diseases: Recent advances and specific challenges in vaccine development. *Vaccines*, 11(2), 470.

- Jaafar, F., Na-Nakorn, U., Srisapoome, P., Amornsakun, T., Duong, T. Y., Gonzales-Plasus, M. M., Hoang, D., & Parhar, I. S. (2021). A current update on the distribution, morphological features, and genetic identity of the Southeast Asian Mahseers, *Tor* species. *Biology*, 10(4), 286.
- Janeway Jr, C. A., & Medzhitov, R. (1998). Introduction: The role of innate immunity in the adaptive immune response. In *Seminars in Immunology*, 10(5), 349-350.
- Janeway, C. (1989). Immunogenecity signals 1, 2, 3... and 0. *Immunology today*, 10(9), 283-286.
- Jault, C., Pichon, L., & Chluba, J. (2004). Toll-like receptor gene family and TIR-domain adapters in *Danio rerio*. *Molecular Immunology*, 40(11), 759-771.
- Jia, D., Wei, L., Guo, W., Zha, R., Bao, M., Chen, Z., Zhao, Y., Ge, C., Zhao, F., Chen, T., Yao, M., Li, J., Wang, H., Gu, J., & He, X. (2011). Genome-wide copy number analyses identified novel cancer genes in hepatocellular carcinoma. *Hepatology*, 54(4), 1227-1236.
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S., Lopez, R., & 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.

- Kaattari, S. L. (1992). Fish B lymphocytes: Defining their form and function. *Annual review of Fish Diseases*, 2, 161-180.
- Kapitonov, V. V., & Jurka, J. (2001). Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences*, 98(15), 8714-8719.
- Kathleen, M. M., Samuel, L., Felecia, C., Reagan, E. L., Kasing, A., Lesley, M., & Toh, S. C. (2016). Antibiotic resistance of diverse bacteria from aquaculture in Borneo. *International Journal of Microbiology*, 2016(1).
- Kaur, M., Jangra, M., Singh, H., Tambat, R., Singh, N., Jachak, S. M., Mishra, S., Sharma, C., Nandanwar, H., & Pinnaka, A. K. (2019). *Pseudomonas koreensis* recovered from raw yak milk synthesizes a β carboline derivative with antimicrobial properties. *Frontiers in Microbiology*, 10, 1728.
- Khanna, R., Kronmiller, B., Maszle, D. R., Coupland, G., Holm, M., Mizuno, T., & Wu, S. H. (2009). The *Arabidopsis* B-box zinc finger family. *The Plant Cell*, 21(11), 3416-3420.
- Kho, C. J. Y., Lau, M. M. L., Chung, H. H., Chew, I. Y. Y., & Gan, H. M. (2023). Whole-genome sequencing of *Pseudomonas koreensis* isolated from diseased *Tor tambroides*. *Current Microbiology*, 80(8), 255.
- Khoo, L. W., Srinivasan, S. S., Henriquez, F. L., & Bal, A. M. (2021). A rare case of mixed infectious keratitis caused by *Pseudomonas koreensis* and *Aspergillus fumigatus*. *Case Reports in Ophthalmology*, 11(3), 600-605.
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357-360.

- Kinne, O. (1980). 14th European marine biology symposium “protection of life in the sea”: Opening address. *Helgoländer Meeresuntersuchungen*, 33, 1-5.
- Kleckner, N. (1981). Transposable elements in prokaryotes. *Annual Review of Genetics*, 15(1), 341-404.
- Kleijer, K. T., Zuko, A., Shimoda, Y., Watanabe, K., & Burbach, J. P. H. (2015). Contactin-5 expression during development and wiring of the thalamocortical system. *Neuroscience*, 310, 106-113.
- Klesius, P. H. (1990). Effect of size and temperature on the quantity of immunoglobulin in channel catfish, *Ictalurus punctatus*. *Veterinary Immunology and Immunopathology*, 24(2), 187-195.
- Kley, R. A., Hellenbroich, Y., Van der Ven, P. F., Fürst, D. O., Huebner, A., Bruchertseifer, V., Peters, S. A., Heyer, C. M., Kirschner, J., Schroder, R., Fischer, D., Muller, K., Tolksdorf, K., Eger, K., Germing, A., Brodherr, T., Reum, C., Walter, M. C., Lochmuller, H., Ketelsen, U. P., & Vorgerd, M. (2007). Clinical and morphological phenotype of the filamin myopathy: A study of 31 German patients. *Brain*, 130(12), 3250-3264.
- Kley, R. A., Serdaroglu-Oflazer, P., Leber, Y., Odgerel, Z., van der Ven, P. F., Olive, M., Ferrer, I., Onipe, A., Mihaylov, M., Bilbao, J. M., Lee, H. S., Hohfeld, J., Djinoivic Carugo, K., Kong, K., Tegenthoff, M., Peter, S. A., Stenzel, W., Vorgerd, M., Goldfarb, L. G., & Fürst, D. O. (2012). Pathophysiology of protein aggregation and extended phenotyping in filaminopathy. *Brain*, 135(9), 2642-2660.
- Klug, A., & Schwabe, J. W. (1995). Zinc fingers. *The FASEB Journal*, 9(8), 597-604.

- Kofler, R., Pandey, R. V., & 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, M. T., Radwan, J., & Babik, W. (2014). Accuracy of allele frequency estimation using pooled RNA-Seq. *Molecular Ecology Resources*, 14(2), 381-392.
- Kongchum, P., Hallerman, E. M., Hulata, G., David, L., & Palti, Y. (2011). Molecular cloning, characterization and expression analysis of TLR9, MyD88 and TRAF6 genes in common carp (*Cyprinus carpio*). *Fish & Shellfish Immunology*, 30(1), 361-371.
- Koser, S. A. (1923). Utilization of the salts of organic acids by the colon-aerogenes group. *Journal of Bacteriology*, 8(5), 493-520.
- Kottelat, M. (2001). *Freshwater fishes of northern Vietnam*. World Bank, Washington, 45-46.
- Kottelat, M. (2013). The fishes of the inland waters of Southeast Asia: A catalogue and core bibliography of the fishes known to occur in freshwaters, mangroves and estuaries. *Raffles Bulletin of Zoology*.
- 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>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547.

- Kuraku, S., Qiu, H., & Meyer, A. (2012). Horizontal transfers of Tc1 elements between teleost fishes and their vertebrate parasites, lampreys. *Genome Biology and Evolution*, 4(9), 929-936.
- Kwon, S. W., Kim, J. S., Park, I. C., Yoon, S. H., Park, D. H., Lim, C. K., & 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.
- Lagerström, M. C., & Schiöth, H. B. (2008). Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nature Reviews Drug discovery*, 7(4), 339-357.
- Lange, S., & Magnadóttir, B. (2003). Spontaneous haemolytic activity of Atlantic halibut (*Hippoglossus hippoglossus* L.) and sea bass (*Dicentrarchus labrax*) serum. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 136(1), 99-106.
- Langefors, Å., Lohm, J., Grahn, M., Andersen, Ø., & Schantz, T. V. (2001). Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1466), 479-485.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359.
- 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.
- Lau, M. M. L., Lim, L. W. K., Chung, H. H., & Gan, H. M. (2021b). The first transcriptome sequencing and data analysis of the Javan mahseer (*Tor tambra*). *Data in Brief*, 39, 107481.
- Lau, M. M. L., Lim, L. W. K., Chung, H. H., & Gan, H. M. (2023). The first draft genome assembly and data analysis of the Malaysian mahseer (*Tor tambroides*). *Aquaculture and Fisheries*, 8(5), 481-491.
- Lau, M. M. L., Lim, L. W. K., Ishak, S. D., Abol-Munafi, A. B., & Chung, H. H. (2021a). A review on the emerging Asian aquaculture fish, the Malaysian mahseer (*Tor tambroides*): Current status and the way forward. In *Proceedings of the Zoological Society*, 1-11. Springer India.
- Laue, B. E., Jiang, Y., Chhabra, S. R., Jacob, S., Stewart, G. S., Hardman, A., Downie, J. A., O’Gara, F., & Williams, P. (2000). The biocontrol strain *Pseudomonas fluorescens* F113 produces the Rhizobium small bacteriocin, N-(3-hydroxy-7-cis-tetradecenoyl) homoserine lactone, via *HdtS*, a putative novel *N*-acylhomoserine lactone synthase. *Microbiology*, 146(10), 2469-2480.

- Lazarovici, P., Primor, N., & Loew, L. M. (1986). Purification and pore-forming activity of two hydrophobic polypeptides from the secretion of the Red Sea Moses sole (*Pardachirus marmoratus*). *Journal of Biological Chemistry*, 261(35), 16704-16713.
- Lefkowitz, R. J. (2004). Historical review: A brief history and personal retrospective of seven transmembrane receptors. *Trends in Pharmacological Sciences*, 25(8), 413-422.
- Lefkowitz, R. J., & Shenoy, S. K. (2005). Transduction of receptor signals by β -arrestins. *Science*, 308(5721), 512-517.
- Lei, Y., Zhou, Y., Price, M., & Song, Z. (2021). Genome-wide characterization of microsatellite DNA in fishes: survey and analysis of their abundance and frequency in genome-specific regions. *BMC Genomics*, 22(1), 1-22.
- Leoni, L., Ciervo, A., Orsi, N., & Visca, P. (1996). Iron-regulated transcription of the *pvdA* gene in *Pseudomonas aeruginosa*: Effect of *Fur* and *PvdS* on promoter activity. *Journal of Bacteriology*, 178(8), 2299-2313.
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, 26(5), 589-595.
- Li, H., Wen, X., Wei, M., Huang, X., Dai, S., Ruan, L., & Yu, Y. (2023). Genome-Wide Identification, Characterization, and Expression Pattern of MYB Gene Family in *Melastoma candidum*. *Horticulturae*, 9(6), 708.
- Li, M., Li, Q., & Wang, Q. L. (2012). Gene expression of TLR5 and TLR5S in channel catfish *Ictalurus punctatus* induced by different pathogens. *Progress in Fishery Sciences*, 33(5), 30-38.

- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., Hvidsten, T. R., Leong, J. S., Minkely, D. R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R. A., Schalburg, K., Rondeau, E. B., Genova, A. D., Samy, J. K. A., Vik, J. O., Vigeland, M. D., Caler, L., Grimholt, U., Jentoft, S., & Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533(7602), 200-205.
- Lihan, S., Jamil, N. A., Jamian, M. A. H., Chiew, T. S., Ajibola, O. O., Justin, S., Benet, F., & Kion, L. N. (2020). Distribution and prevalence of antibiotic resistant bacteria in fish farms in East Malaysia. *Malaysian Journal of Microbiology*, 16(4).
- Lim, L. W. K., Chung, H. H., Hussain, H., & Gan, H. M. (2021c). Genome survey of sago palm (*Metroxylon sagu* Rottboll). *Plant Gene*, 28, 100341.
- Lim, L. W. K., Chung, H. H., Ishak, S. D., & Waiho, K. (2021b). Zebrafish (*Danio rerio*) ecotoxicological ABCB4, ABCC1 and ABCG2a gene promoters depict spatiotemporal xenobiotic multidrug resistance properties against environmental pollutants. *Gene Reports*, 23, 101110.
- Lim, L. W. K., Chung, H. H., Lau, M. M. L., Aziz, F., & Gan, H. M. (2021a). Improving the phylogenetic resolution of Malaysian and Javan mahseer (*Cyprinidae*), *Tor tambroides* and *Tor tambra*: Whole mitogenomes sequencing, phylogeny and potential mitogenome markers. *Gene*, 791, 145708.
- Lim, L. W. K., Lau, M. M. L., Chung, H. H., Hussain, H., & Gan, H. M. (2022). First high-quality genome assembly data of sago palm (*Metroxylon sagu* Rottboll). *Data in Brief*, 40, 107800.

- Lim, L. W. K., Tan, H. Y., Aminan, A. W., Jumaan, A. Q., Moktar, M. Z., Tan, S. Y., Balinu, C. P., Robert, A. V., Chung, H. H., & Sulaiman, B. (2018). Phylogenetic and expression of ATP-binding cassette transporter genes in *Rasbora sarawakensis*. *Pertanika Journal of Tropical Agricultural Science*, 41(3), 1341-1354.
- Lin, G., Thevasagayam, N. M., Wan, Z. Y., Ye, B. Q., & Yue, G. H. (2019). Transcriptome analysis identified genes for growth and omega-3/-6 ratio in saline tilapia. *Frontiers in Genetics*, 10, 244.
- Lionel, A. C., Crosbie, J., Barbosa, N., Goodale, T., Thiruvahindrapuram, B., Rickaby, J., Gazzellone, M., Carson, A. R., Howe, J. L., Wang, Z., Wei, J., Stewart, A. F. R., Roberts, R., McPherson, R., Fiebig, A., Franke, A., Schreiber, S., Zwaigenbaum, L., Fernandez, B. A., Roberts, W., Arnold, P. D., Szatmari, P., Marshall, C. R., Schachar, R., & 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.
- Liu, Y., Rzeszutek, E., Van Der Voort, M., Wu, C. H., Thoen, E., Skaar, I., Bulone, V., Dorrestein, P. C., Raaijmakers, J. M., & De Bruijn, I. (2015). Diversity of aquatic *Pseudomonas* species and their activity against the fish pathogenic oomycete *Saprolegnia*. *PLOS One*, 10(8), e0136241.
- Lopes, L. D., Pereira e Silva, M. D. C., Weisberg, A. J., Davis, E. W., Yan, Q., Varize, C. D. S., Wu, C., Chang, J. H., Loper, J. E., & Andreote, F. D. (2018). Genome variations between rhizosphere and bulk soil ecotypes of a *Pseudomonas koreensis* population. *Environmental Microbiology*, 20(12), 4401-4414.

- Lopez, J. R., Dieguez, A. L., Doce, A., De la Roca, E., De la Herran, R., Navas, J. I., Toranzo, A. E., & Romalde, J. L. (2012). *Pseudomonas baetica* sp. nov., a fish pathogen isolated from wedge sole, *Dicologlossa cuneata* (Moreau). *International Journal of Systematic and Evolutionary Microbiology*, 62(Pt_4), 874-882.
- Lozano, G. L., Bravo, J. I., & Handelsman, J. (2017). Draft genome sequence of *Pseudomonas koreensis* CI12, a *Bacillus cereus* “hitchhiker” from the soybean rhizosphere. *Genome Announcements*, 5(26), e00570-17.
- Lu, G., & Luo, M. (2020). Genomes of major fishes in world fisheries and aquaculture: Status, application and perspective. *Aquaculture and Fisheries*, 5(4), 163-173.
- Lu, X., Kong, J., Meng, X., Cao, B., Luo, K., Dai, P., & Luan, S. (2018). Identification of SNP markers associated with tolerance to ammonia toxicity by selective genotyping from *de novo* assembled transcriptome in *Litopenaeus vannamei*. *Fish & Shellfish Immunology*, 73, 158-166.
- Lu, Y. C., Yeh, W. C., & Ohashi, P. S. (2008). LPS/TLR4 signal transduction pathway. *Cytokine*, 42(2), 145-151.
- Lumsden, J. S., Ostland, V. E., MacPhee, D. D., & Ferguson, H. W. (1995). Production of gill associated and serum antibody by rainbow trout (*Oncorhynchus mykiss*) following immersion immunization with acetone-killed *Flavobacterium branchiophilum* and the relationship to protection from experimental challenge. *Fish & Shellfish Immunology*, 5(2), 151-165.
- Lund, V., Arnesen, J. A., Mikkelsen, H., Gravningen, K., Brown, L., & Schrøder, M. B. (2008). Atypical furunculosis vaccines for Atlantic cod (*Gadus morhua*); vaccine efficacy and antibody responses. *Vaccine*, 26(52), 6791-6799.

- 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.
- Ma, D., Ma, A., Huang, Z., Wang, G., Wang, T., Xia, D., & Ma, B. (2016). Transcriptome analysis for identification of genes related to gonad differentiation, growth, immune response and marker discovery in the turbot (*Scophthalmus maximus*). *PLOS One*, 11(2), e0149414.
- MacGowan, A., & Macnaughton, E. (2017). Antibiotic resistance. *Medicine*, 45(10), 622-628.
- Madigan, M. T., Martinko, J. M., & Parker, J. (2000). *Brock biology of microorganisms* (Vol. 11). Upper Saddle River, NJ: Prentice hall.
- Magnadóttir, B. (2010). Immunological control of fish diseases. *Marine biotechnology*, 12, 361-379.
- Magnadóttir, B., Bambir, S. H., Gudmundsdóttir, B. K., Pilström, L., & Helgason, S. (2002). Atypical *Aeromonas salmonicida* infection in naturally and experimentally infected cod, *Gadus morhua* L. *Journal of Fish Diseases*, 25(10), 583-597.
- Magnadóttir, B., Lange, S., Gudmundsdottir, S., Bøggwald, J., & Dalmo, R. A. (2005). Ontogeny of humoral immune parameters in fish. *Fish & Shellfish Immunology*, 19(5), 429-439.
- Mahajan-Miklos, S., Tan, M. W., Rahme, L. G., & Ausubel, F. M. (1999). Molecular mechanisms of bacterial virulence elucidated using a *Pseudomonas aeruginosa* *Caenorhabditis elegans* pathogenesis model. *Cell*, 96(1), 47-56.

- Malmstrøm, M., Matschiner, M., Tørresen, O. K., Star, B., Snipen, L. G., Hansen, T. F., ... & Jentoft, S. (2016). Evolution of the immune system influences speciation rates in teleost fishes. *Nature Genetics*, 48(10), 1204-1210.
- Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, 38(10), 4647-4654.
- 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.
- Marburger, S., Alexandrou, M. A., Taggart, J. B., Creer, S., Carvalho, G., Oliveira, C., & Taylor, M. I. (2018). Whole genome duplication and transposable element proliferation drive genome expansion in *Corydoradinae* catfishes. *Proceedings of the Royal Society B: Biological Sciences*, 285(1872), 20172732.
- Marc, C., Vrignaud, B., Levieux, K., Robine, A., Guen, C. G. L., & Launay, E. (2016). Inappropriate prescription of antibiotics in pediatric practice: Analysis of the prescriptions in primary care. *Journal of Child Health Care*, 20(4), 530-536.
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of kmers. *Bioinformatics Methods and Protocols*, 27(6), 764-770.
- Masri, L., Branca, A., Sheppard, A. E., Papkou, A., Laehnemann, D., Guenther, P. S., Prahl, S., Saebelfeld, M., Hollensteiner, J., Liesegang, H., Brzuszkiewicz, E., Daniel, R., Michiels, N. K., Schulte, R. D., Kurtz, J., Rosenstiel, R., Telschow, A., Bornberg

- Bauer, E., & Schulenburg, H. (2015). Host–pathogen coevolution: The selective advantage of *Bacillus thuringiensis* virulence and its cry toxin genes. *PLOS biology*, 13(6), e1002169.
- Matsunaga, T. (1998). Did the first adaptive immunity evolve in the gut of ancient jawed fish?. *Cytogenetics and Cell Genetics*, 80(1-4), 138-141.
- Matyar, F., Kaya, A., & Dinçer, S. (2008). Antibacterial agents and heavy metal resistance in Gram-negative bacteria isolated from seawater, shrimp and sediment in Iskenderun Bay, Turkey. *Science of the Total Environment*, 407(1), 279-285.
- Maule, A. G., Schrock, R., Slater, C., Fitzpatrick, M. S., & Schreck, C. B. (1996). Immune and endocrine responses of adult chinook salmon during freshwater immigration and sexual maturation. *Fish & Shellfish Immunology*, 6(3), 221-233.
- Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sanstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kuttyavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., Bates, D., Hansen, R. S., Neph, S., Sabo, P. J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S. R., Kaul, R., & Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease associated variation in regulatory DNA. *Science*, 337(6099), 1190-1195.
- Meijer, A. H., Krens, S. G., Rodriguez, I. A. M., He, S., Bitter, W., Snaar-Jagalska, B. E., & Spaik, H. P. (2004). Expression analysis of the Toll-like receptor and TIR domain adaptor families of zebrafish. *Molecular Immunology*, 40(11), 773-783.
- Mena, K. D., & Gerba, C. P. (2009). Risk assessment of *Pseudomonas aeruginosa* in water. *Reviews of Environmental Contamination and Toxicology*, 201, 71-115.

- Migliori, G. B., Tiberi, S., Zumla, A., Petersen, E., Chakaya, J. M., Wejse, C., Torrico, M. M., Duarte, R., Alffenaar, J. W., Schaaf, H. S., Marais, B. J., Cirillo, D. M., Alagna, R., Rendon, A., Pontali, E., Piubello, A., Figueroa, J., Ferlazzo, G., Garcia-Basteiro, A., Centis, R., Visca, D., D'Ambrosio, L., Sotgiu, G., & Zellweger, J. P. (2020). MDR/XDR-TB management of patients and contacts: Challenges facing the new decade. The 2020 clinical update by the Global Tuberculosis Network. *International Journal of Infectious Diseases*, 92, S15-S25.
- Miller, L. C., & Tainter, M. (1944). Estimation of the ED50 and its error by means of logarithmic-probit graph paper. *Proceedings of the society for Experimental Biology and Medicine*, 57(2), 261-264.
- Mittal, K. R., Lalonde, G., Leblanc, D., Olivier, G., & Lallier, R. (1980). *Aeromonas hydrophila* in rainbow trout: Relation between virulence and surface characteristics. *Canadian Journal of Microbiology*, 26(12), 1501-1503.
- Moen, T., Baranski, M., Sonesson, A. K., & Kjøglum, S. (2009). Confirmation and fine mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): Population-level associations between markers and trait. *BMC Genomics*, 10, 1-14.
- 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.
- Moore, H., Greenwell, P. W., Liu, C. P., Arnheim, N., & Petes, T. D. (1999). Triplet repeats form secondary structures that escape DNA repair in yeast. *Proceedings of the National Academy of Sciences*, 96(4), 1504-1509.

- Moss, S. M., Moss, D. R., Arce, S. M., Lightner, D. V., & Lotz, J. M. (2012). The role of selective breeding and biosecurity in the prevention of disease in penaeid shrimp aquaculture. *Journal of Invertebrate Pathology*, 110(2), 247-250.
- Mousseau, T. A., & Fox, C. W. (1998). The adaptive significance of maternal effects. *Trends in Ecology & Evolution*, 13(10), 403-407.
- Munita, J. M., & Arias, C. A. (2016). Mechanisms of antibiotic resistance. *Virulence mechanisms of bacterial pathogens*, 481-511.
- Nadjar-Boger, E., & Funkenstein, B. (2011). Myostatin-2 gene structure and polymorphism of the promoter and first intron in the marine fish *Sparus aurata*: Evidence for DNA duplications and/or translocations. *BMC Genetics*, 12(1), 1-19.
- Nagae, M., Fuda, H., Hara, A., Kawamura, H., & Yamauchi, K. (1993). Changes in serum immunoglobulin M (IgM) concentrations during early development of chum salmon (*Oncorhynchus keta*) as determined by sensitive ELISA technique. *Comparative Biochemistry and Physiology Part A: Physiology*, 106(1), 69-74.
- Nakabayashi, K., Komaki, G., Tajima, A., Ando, T., Ishikawa, M., Nomoto, J., Hata, K., Oka, A., Inoko, H., Sasazuki, T., JGRED., & 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.
- Nakanishi, T. (1986). Seasonal changes in the humoral immune response and the lymphoid tissues of the marine teleost, *Sebastiscus marmoratus*. *Veterinary Immunology and Immunopathology*, 12(1-4), 213-221.

- Nakanishi, T., Toda, H., Shibasaki, Y., & Somamoto, T. (2011). Cytotoxic T cells in teleost fish. *Developmental & Comparative Immunology*, 35(12), 1317-1323.
- Nandi, D., Tahiliani, P., Kumar, A., & Chandu, D. (2006). The ubiquitin-proteasome system. *Journal of Biosciences*, 31, 137-155.
- Neefjes, J., Jongsma, M. L., Paul, P., & Bakke, O. (2011). Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, 11(12), 823-836.
- Neilands, J. B. (1981). Iron absorption and transport in microorganisms. *Annual Review of Nutrition*, 1(1), 27-46.
- Nepal, G., & Bhatta, S. (2018). Self-medication with antibiotics in WHO Southeast Asian Region: A systematic review. *Cureus*, 10(4).
- Ng, C. K. (2004). Kings of the river—Mahseer in Malaysia and the region. Kuala Lumpur, Malaysia: Inter Sea Fishery.
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268-274.
- Nie, L., Cai, S. Y., Shao, J. Z., & Chen, J. (2018). Toll-like receptors, associated biological roles, and signaling networks in non-mammals. *Frontiers in Immunology*, 9, 1523.
- Nikpay, M., Šeda, O., Tremblay, J., Petrovich, M., Gaudet, D., Kotchen, T. A., Cowley, A. W., & 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. *Hypertension Research*, 35(6), 585-591.

- Njeim, R., & Eid, A. (2018). Alpha - 2c adrenergic receptor promotes the malignant phenotype of colon cancer cells. *The FASEB Journal*, 32, 695-5.
- Novotny, L., Dvorska, L., Lorencova, A., Beran, V., & Pavlik, I. (2004). Fish: A potential source of bacterial pathogens for human beings. *Veterinarni Medicina*, 49(9), 343-358.
- O'Malley, M. R., Chien, C. F., Peck, S. C., Lin, N. C., & Anderson, J. C. (2020). A revised model for the role of GacS/GacA in regulating type III secretion by *Pseudomonas syringae* pv. tomato DC3000. *Molecular Plant Pathology*, 21(1), 139-144.
- Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., & 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.
- Ogawa, J., Kaneko, H., Masuda, T., Nagata, S., Hosoya, H., & Watanabe, K. (1996). Novel neural adhesion molecules in the Contactin/F3 subgroup of the immunoglobulin superfamily: Isolation and characterization of cDNAs from rat brain. *Neuroscience Letters*, 218(3), 173-176.
- Oguro-Ando, A., Zuko, A., Kleijer, K. T., & Burbach, J. P. H. (2017). A current view on contactin-4,-5, and-6: Implications in neurodevelopmental disorders. *Molecular and Cellular Neuroscience*, 81, 72-83.
- Oh, W. T., Kim, J. H., Jun, J. W., Giri, S. S., Yun, S., Kim, H. J., Kim, S. G., Kim, S. W., Han, S. J., Kwon, J., & Park, S. C. (2019). Genetic characterization and pathological analysis of a novel bacterial pathogen, *Pseudomonas tructae*, in rainbow trout (*Oncorhynchus mykiss*). *Microorganisms*, 7(10), 432.

- Oliver, J. L., & Marín, A. (1996). A relationship between GC content and coding-sequence length. *Journal of Molecular Evolution*, 43, 216-223.
- Oshiumi, H., Tsujita, T., Shida, K., Matsumoto, M., Ikeo, K., & Seya, T. (2003). Prediction of the prototype of the human Toll-like receptor gene family from the pufferfish, *Fugu rubripes*, genome. *Immunogenetics*, 54(11), 791-800.
- Overturf, K., Sakhrani, D., & Devlin, R. H. (2010). Expression profile for metabolic and growth-related genes in domesticated and transgenic coho salmon (*Oncorhynchus kisutch*) modified for increased growth hormone production. *Aquaculture*, 307(1-2), 111-122.
- Palleroni, N. J. (1983). Introduction to the Family *Pseudomonadaceae*. In: *The procaryotes, a handbook on habitat, isolation and identification of bacteria* (pp. 655-718). Springer Verlag: New York.
- Palti, Y. (2011). Toll-like receptors in bony fish: From genomics to function. *Developmental & Comparative Immunology*, 35(12), 1263-1272.
- Pękala-Safińska, A. (2018). Contemporary threats of bacterial infections in freshwater fish. *Journal of Veterinary Research*, 62(3), 261.
- Pérez-Casanova, J. C., Rise, M. L., Dixon, B., Afonso, L. O. B., Hall, J. R., Johnson, S. C., & Gamperl, A. K. (2008). The immune and stress responses of Atlantic cod to long-term increases in water temperature. *Fish & Shellfish Immunology*, 24(5), 600-609.
- Pessi, G., & Haas, D. (2000). Transcriptional control of the hydrogen cyanide biosynthetic genes *hcnABC* by the anaerobic regulator ANR and the quorum-sensing regulators *LasR* and *RhlR* in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 182(24), 6940-6949.

- Peymani, A., Naserpour-Farivar, T., Zare, E., & Azarhoosh, K. H. (2017). Distribution of bla_{TEM}, bla_{SHV}, and bla_{CTX-M} genes among ESBL-producing *P. aeruginosa* isolated from Qazvin and Tehran hospitals, Iran. *Journal of Preventive Medicine and Hygiene*, 58(2), E155.
- Picchietti, S., Terribili, F. R., Mastrolia, L., Scapigliati, G., & Abelli, L. (1997). Expression of lymphocyte antigenic determinants in developing gut-associated lymphoid tissue of the sea bass *Dicentrarchus labrax* (L.). *Anatomy and Embryology*, 196, 457-463.
- Pickett, B. D., Miller, J. B., & Ridge, P. G. (2017). Kmer-SSR: A fast and exhaustive SSR search algorithm. *Bioinformatics*, 33(24), 3922-3928.
- Pilström, L., Warr, G. W., & Strömberg, S. (2005). Why is the antibody response of Atlantic cod so poor? The search for a genetic explanation. *Fisheries Science*, 71, 961-971.
- Pinder, A. C., Britton, J. R., Harrison, A. J., Nautiyal, P., Bower, S. D., Cooke, S. J., Lockett, S., Everard, M., Katwate, U., Ranjeet, K., Walton, S., Danylchuk, A. J., Dahanukar, N., Raghavan, R. (2019). Mahseer (*Tor* spp.) fishes of the world: Status, challenges and opportunities for conservation. *Reviews in Fish Biology and Fisheries*, 29(2), 417-452.
- Pineda-Cirera, L., Cabana-Dominguez, J., Lee, P. H., Fernandez-Castillo, N., & Cormand, B. (2022). Identification of genetic variants influencing methylation in brain with pleiotropic effects on psychiatric disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 113, 110454.
- Plasterk, R. H., Izsvák, Z., & Ivics, Z. (1999). Resident aliens: The Tc1/mariner superfamily of transposable elements. *Trends in Genetics*, 15(8), 326-332.

- Pontarotti, P. (Ed.). (2017). *Evolutionary biology: Self/nonsself evolution, species and complex traits evolution, methods and concepts*. Springer.
- Powell, K. (1993). Is biological control the answer for sustainable agriculture?. *Mycologist*, 7(2), 75-78.
- Praveena, S. M., Shaifuddin, S. N. M., Sukiman, S., Nasir, F. A. M., Hanafi, Z., Kamarudin, N., Ismail, T. H. T., & Aris, A. Z. (2018). Pharmaceuticals residues in selected tropical surface water bodies from Selangor (Malaysia): Occurrence and potential risk assessments. *Science of the Total Environment*, 642, 230-240.
- Quillet, E., Dorson, M., Le Guillou, S., Benmansour, A., & Boudinot, P. (2007). Wide range of susceptibility to rhabdoviruses in homozygous clones of rainbow trout. *Fish & Shellfish Immunology*, 22(5), 510-519.
- Rahayu, D. A., & Nugroho, E. D. (2015). DNA barcode dan haplotype network ikan local dari telaga banyu biru kabupaten pasuruan. In *Seminar Nasional Ikan Ke-8; Masyarakat Iktiologi Indonesia: Bogor, Indonesia* (Vol. 2, pp. 67-75).
- Raida, M. K., & Buchmann, K. (2009). Innate immune response in rainbow trout (*Oncorhynchus mykiss*) against primary and secondary infections with *Yersinia ruckeri* O1. *Developmental & Comparative Immunology*, 33(1), 35-45.
- Ramirez-Gomez, F., Greene, W., Rego, K., Hansen, J. D., Costa, G., Kataria, P., & Bromage, E. S. (2012). Discovery and characterization of secretory IgD in rainbow trout: Secretory IgD is produced through a novel splicing mechanism. *The Journal of Immunology*, 188(3), 1341-1349.
- Randelli, E., Buonocore, F., & Scapigliati, G. (2008). Cell markers and determinants in fish immunology. *Fish & Shellfish Immunology*, 25(4), 326-340.

- Rauta, P. R., Nayak, B., & Das, S. (2012). Immune system and immune responses in fish and their role in comparative immunity study: A model for higher organisms. *Immunology Letters*, 148(1), 23-33.
- Razquin, B. E., Castillo, A., Lopez-Fierro, P., Alvarez, F., Zapata, A., & Villena, A. J. (1990). Ontogeny of IgM-producing cells in the lymphoid organs of rainbow trout, *Salmo gairdneri* Richardson: An immuno- and enzyme-histochemical study. *Journal of Fish Biology*, 36(2), 159-173.
- Reboud, E., Basso, P., Maillard, A. P., Huber, P., & Attrée, I. (2017). Exolysin shapes the virulence of *Pseudomonas aeruginosa* clonal outliers. *Toxins*, 9(11), 364-376.
- Renaville, B., Bacciu, N., Lanzoni, M., Mossa, F., & Piasentier, E. (2018). Association of single nucleotide polymorphisms in fat metabolism candidate genes with fatty acid profiles of muscle and subcutaneous fat in heavy pigs. *Meat Science*, 139, 220–227.
- Ribeiro da Cunha, B., Fonseca, L. P., & Calado, C. R. (2019). Antibiotic discovery: Where have we come from, where do we go?. *Antibiotics*, 8(2), 45.
- Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O. J., Samaha, R. R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J. Z., Ghandehari, D., Sherman, B. K., & L. Yu, G. (2000). *Arabidopsis* transcription factors: Genome-wide comparative analysis among eukaryotes. *Science*, 290(5499), 2105-2110.
- Rivero, G., Martín-Guerrero, I., de Prado, E., Gabilondo, A. M., Callado, L. F., García Sevilla, J. A., Garcia-Orad, A., & Meana, J. J. (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.
- Roach, J. C., Glusman, G., Rowen, L., Kaur, A., Purcell, M. K., Smith, K. D., Hood, L. E., & Aderem, A. (2005). The evolution of vertebrate Toll-like receptors. *Proceedings of the National Academy of Sciences*, 102(27), 9577-9582.
- Roberts, T. R. & M. Z. Khaironizam (2008). Trophic polymorphism in the Malaysian fish *Neolissochilus soroides* and other old world barbs (Teleostei, Cyprinidae). *Natural History Bulletin of the Siam Society*, 56, 25–53.
- Roberts, T. R. (1993). The freshwater fishes of Java, as observed by Kuhl and van Hasselt in 1820-23. *Zoologische Verhandelingen*, 285(1), 1-94.
- Roberts, T. R. (1999). Fishes of the cyprinid genus *Tor* in the Nam Theun watershed (Mekong basin) of Laos, with description of a new species. *Raffles Bulletin of Zoology*, 47, 225-236.
- Robertsen, B. (2006). The interferon system of teleost fish. *Fish & Shellfish Immunology*, 20(2), 172-191.
- Røed, K. H., Fjalested, K. T., & Strømsheim, A. (1993). Genetic variation in non-specific immune parameters in Atlantic salmon (*Salmo salar*). *Aquaculture*, 111(1-4), 294.
- Rohani, M. F., Islam, S. M., Hossain, M. K., Ferdous, Z., Siddik, M. A., Nuruzzaman, M., Padeniya, Y., Brown, C., & 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.
- Rolfe, M. D., Rice, C. J., Lucchini, S., Pin, C., Thompson, A., Cameron, A. D., Alston, M., Stringer, M. F., Betts, R. P., Baranyi, J., Peck, M. W., & Hinton, J. C. (2012). Lag

- phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *Journal of Bacteriology*, 194(3), 686-701.
- Rolff, J. (2007). Why did the acquired immune system of vertebrates evolve?. *Developmental & Comparative Immunology*, 31(5), 476-482.
- Rombel, I. T., McMorran, B. J., & Lamont, I. L. (1995). Identification of a DNA sequence motif required for expression of iron-regulated genes in *Pseudomonads*. *Molecular and General Genetics MGG*, 246, 519-528.
- Sahoo, B. R. (2020). Structure of fish Toll-like receptors (TLR) and NOD-like receptors (NLR). *International Journal of Biological Macromolecules*, 161, 1602-1617.
- Sakai, D. K. (1992). Repertoire of complement in immunological defense mechanisms of fish. *Annual Review of Fish Diseases*, 2, 223-247.
- Salinas, I., & Magadán, S. (2017). Omics in fish mucosal immunity. *Developmental & Comparative Immunology*, 75, 99-108.
- Sawayama, E., Tanizawa, S., Kitamura, S. I., Nakayama, K., Ohta, K., Ozaki, A., & Takagi, M. (2017). Identification of quantitative trait loci for resistance to RSIVD in red sea bream (*Pagrus major*). *Marine Biotechnology*, 19, 601-613.
- Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., Wang, J., Williams, R., Trawick, B. W., Pruitt, K. D., & Sherry, S. T. (2021). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 49(D1), D10.

- Scales, B. S., Dickson, R. P., LiPuma, J. J., & Huffnagle, G. B. (2014). Microbiology, genomics, and clinical significance of the *Pseudomonas fluorescens* species complex, an unappreciated colonizer of humans. *Clinical Microbiology Reviews*, 27(4), 927-948.
- Scapigliati, G., Fausto, A. M., & Picchietti, S. (2018). Fish lymphocytes: An evolutionary equivalent of mammalian innate-like lymphocytes?. *Frontiers in Immunology*, 9, 971.
- Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma*, 109(6), 365-371.
- Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals —mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, 15(11), 749-763.
- Schrader, L., Kim, J. W., Ence, D., Zimin, A., Klein, A., Wyschetzki, K., Weichselgartner, T., Kemena, C., Stokl, J., Schultner, E., Wurm, Y., Smith, C. D., Yandell, M., Heinze, J., Gadau, J., & Oettler, J. (2014). Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications*, 5(1), 1-10.
- Schrøder, M. B., Villena, A. J., & Jørgensen, T. Ø. (1998). Ontogeny of lymphoid organs and immunoglobulin producing cells in Atlantic cod (*Gadus morhua* L.). *Developmental & Comparative Immunology*, 22(5-6), 507-517.
- Sciuto, S., Colli, L., Fabris, A., Pastorino, P., Stoppani, N., Esposito, G., Prearo, M., Esposito, G., Ajmone-Marsan, P., Acutis, P. L., & Colussi, S. (2022). What can genetics do for the control of infectious diseases in aquaculture?. *Animals*, 12(17), 2176.

- SEAFDEC. (2017). Southeast Asian State of Fisheries and Aquaculture 2017. Southeast Asian Fisheries Development Center, 167.
- Secombes, C. J., & Fletcher, T. C. (1992). The role of phagocytes in the protective mechanisms of fish. *Annual Review of Fish Diseases*, 2, 53-71.
- Secombes, C. J., Hardie, L. J., & Daniels, G. (1996). Cytokines in fish: An update. *Fish & Shellfish Immunology*, 6(4), 291-304.
- Senderovich, Y., Ken-Dror, S., Vainblat, I., Blau, D., Izhaki, I., & Halpern, M. (2012). A molecular study on the prevalence and virulence potential of *Aeromonas* spp. recovered from patients suffering from diarrhea in Israel. *PLOS One*, 7(2), e30070.
- Shahi, N., & Malik, S. K. (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.
- Shao, F., Han, M., & Peng, Z. (2019). Evolution and diversity of transposable elements in fish genomes. *Scientific Reports*, 9(1), 15399.
- Shaul, O. (2017). How introns enhance gene expression. *The International Journal of Biochemistry & Cell Biology*, 91, 145-155.
- Short, K. M., & Cox, T. C. (2006). Subclassification of the RBCC/TRIM superfamily reveals a novel motif necessary for microtubule binding. *Journal of Biological Chemistry*, 281(13), 8970-8980.
- Siraj, S.S., Esa, Y.B., Keong, B.P., & Daud, S.K., (2007). Genetic characterization of the two colour-type of Kelah. *Malaysian Applied Biology*, 36(1), 23–29.

- Sitjà-Bobadilla, A., Redondo, M. J., Bermúdez, R., Palenzuela, O., Ferreiro, I., Riaza, A., Quiroga, I., Nieto, J. M., & Alvarez-Pellitero, P. (2006). Innate and adaptive immune responses of turbot, *Scophthalmus maximus* (L.), following experimental infection with *Enteromyxum scophthalmi* (Myxosporea: Myxozoa). *Fish & Shellfish Immunology*, 21(5), 485-500.
- Sreeja, M. K., Gowrishankar, N. L., Adisha, S., & Divya, K. C. (2017). Antibiotic resistance-reasons and the most common resistant pathogens-A review. *Research Journal of Pharmacy and Technology*, 10(6), 1886-1890.
- 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, T. P., Condeelis, J., Cooley, L., Hartwig, J. H., Noegel, A., Schleicher, M., & Shapiro, S. S. (2001). Filamins as integrators of cell mechanics and signalling. *Nature Reviews Molecular Cell Biology*, 2(2), 138-145.
- Subramaniam, G., & Girish, M. (2020). Antibiotic resistance—A cause for reemergence of infections. *The Indian Journal of Pediatrics*, 87(11), 937-944.
- Sulis, G., Daniels, B., Kwan, A., Gandra, S., Daftary, A., Das, J., & Pai, M. (2020). Antibiotic overuse in the primary health care setting: A secondary data analysis of standardised patient studies from India, China and Kenya. *BMJ Global Health*, 5(9), e003393.
- Sup Lee, J., Hanford, M. G., Genova, J. L., & Farber, R. (1999). Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. *Human Molecular Genetics*, 8(13), 2567-2572.

- Swain, P., & Nayak, S. K. (2009). Role of maternally derived immunity in fish. *Fish & Shellfish Immunology*, 27(2), 89-99.
- Szalai, A. J., Bly, J. E., & Clem, L. W. (1994). Changes in serum concentrations of channel catfish (*Ictalurus punctatus* Rafinesque) phosphorylcholine-reactive protein (PRP) in response to inflammatory agents, low temperature-shock and infection by the fungus *Saprolegnia* sp. *Fish & Shellfish Immunology*, 4(5), 323-336.
- Tahirov, T. H., 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 Crystallographica Section D: Biological Crystallography*, 57(11), 1655-1658.
- Takemura, A. (1993). Changes in an immunoglobulin M (IgM)-like protein during larval stages in tilapia, *Oreochromis mossambicus*. *Aquaculture*, 115(3-4), 233-241.
- Takemura, A., & Takano, K. (1997). Transfer of maternally-derived immunoglobulin (IgM) to larvae in tilapia, *Oreochromis mossambicus*. *Fish & shellfish immunology*, 7(6), 355-363.
- Tamminen, M., Karkman, A., Lohmus, A., Muziasari, W. I., Takasu, H., Wada, S., Suzuki, S., & Virta, M. (2011). Tetracycline resistance genes persist at aquaculture farms in the absence of selection pressure. *Environmental Science & Technology*, 45(2), 386-391.
- Tan, M. H., Austin, C. M., Hammer, M. P., Lee, Y. P., Croft, L. J., & Gan, H. M. (2018). Finding Nemo: Hybrid assembly with Oxford Nanopore and Illumina reads greatly

- improves the clownfish (*Amphiprion ocellaris*) genome assembly. *Gigascience*, 7(3), 1-6.
- Tan, Y. P., Lin, Q., Wang, X. H., Joshi, S., Hew, C. L., & Leung, K. Y. (2002). Comparative proteomic analysis of extracellular proteins of *Edwardsiella tarda*. *Infection and Immunity*, 70(11), 6475-6480.
- Tang, H. (2007). Genome assembly, rearrangement, and repeats. *Chemical Reviews*, 107(8), 3391-3406.
- Tóth, G., Gáspári, Z., & Jurka, J. (2000). Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research*, 10(7), 967-981.
- Tsujita, T., Tsukada, H., Nakao, M., Oshiumi, H., Matsumoto, M., & Seya, T. (2004). Sensing bacterial flagellin by membrane and soluble orthologs of Toll-like receptor 5 in rainbow trout (*Onchorhynchus mikiss*). *Journal of Biological Chemistry*, 279(47), 48588-48597.
- Tu, H. T., Silvestre, F., Scippo, M. L., Thome, J. P., Phuong, N. T., & Kestemont, P. (2009). Acetylcholinesterase activity as a biomarker of exposure to antibiotics and pesticides in the black tiger shrimp (*Penaeus monodon*). *Ecotoxicology and Environmental Safety*, 72(5), 1463-1470.
- Uddin, T. M., Chakraborty, A. J., Khusro, A., Zidan, B. R. M., Mitra, S., Emran, T. B., Dhama, K., Ripon, M. K. H., Gajdacs, M., Sahibzada, M. U. K., Hossain, M. J., & Koirala, N. (2021). Antibiotic resistance in microbes: History, mechanisms, therapeutic strategies and future prospects. *Journal of Infection and Public Health*, 14(12), 1750-1766.

- Uribe, C., Folch, H., Enríquez, R., & Moran, G. J. V. M. (2011). Innate and adaptive immunity in teleost fish: A review. *Veterinarni medicina*, 56(10), 486-503.
- Vagner, M., & Santigosa, E. (2011). Characterization and modulation of gene expression and enzymatic activity of delta-6 desaturase in teleosts: A review. *Aquaculture*, 315(1-2), 131-143.
- Valentini, M., Gonzalez, D., Mavridou, D. A., & Filloux, A. (2018). Lifestyle transitions and adaptive pathogenesis of *Pseudomonas aeruginosa*. *Current Opinion in Microbiology*, 41, 15-20.
- Van Muiswinkel, W. B., Wiegertjes, G. F., & Stet, R. J. (1999). The influence of environmental and genetic factors on the disease resistance of fish. *Aquaculture*, 172(1-2), 103-110.
- Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: Causes and threats. *Pharmacy and Therapeutics*, 40(4), 277.
- Verjan, N., Hirono, I., & Aoki, T. (2005). Genetic loci of major antigenic protein genes of *Edwardsiella tarda*. *Applied and Environmental Microbiology*, 71(9), 5654-5658.
- 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. *Genetics Selection Evolution*, 34(3), 275-305.
- Vij, S., Kuhl, H., Kuznetsova, I. S., Komissarov, A., Yurchenko, A. A., Van Heusden, P., Singh, S., Thevasagayam, N. M., Prakki, S. R. S., Purushothaman, K., Saju, J. M., Jiang, J., & Orbán, L. (2016). Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *PLoS Genetics*, 12(4), e1005954.

- Volff, J. N. (2005). Genome evolution and biodiversity in teleost fish. *Heredity*, 94(3), 280-294.
- Von Wintersdorff, C. J., Penders, J., Van Niekerk, J. M., Mills, N. D., Majumder, S., Van Alphen, L. B., Savelkoul, P. H. M., & Wolffs, P. F. (2016). Dissemination of antimicrobial resistance in microbial ecosystems through horizontal gene transfer. *Frontiers in Microbiology*, 7, 173.
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, 33(14), 2202-2204.
- Waiho, K., Fazhan, H., Ishak, S. D., Kasan, N. A., Liew, H. J., Norainy, M. H., & Ikhwanuddin, M. (2020). Potential impacts of COVID-19 on the aquaculture sector of Malaysia and its coping strategies. *Aquaculture Reports*, 18, 100450.
- Wakabayashi, H., Sawada, K., Ninomiya, K., & Nishimori, E. (1996). Bacterial hemorrhagic ascites of ayu caused by *Pseudomonas* sp. *Fish Pathology*, 31(4), 239-240.
- Walton, S. E., Gan, H. M., Raghavan, R., Pinder, A. C., Ahmad, A. (2017). Disentangling] the taxonomy of the mahseers (*Tor* spp.) of Malaysia: An integrated approach using morphology, genetics and historical records. *Reviews in Fisheries Science and Aquaculture*, 25(3), 171–183.
- Wang, J., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, S., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczski, C. J., & Marchler-Bauer, A. (2023). The conserved domain database in 2023. *Nucleic Acids Research*, 51(D1), D384-D388.

- Wang, R. H., Li, W., Fan, Y. D., Liu, Q. L., Zeng, L. B., & Xiao, T. Y. (2017). Tlr22 structure and expression characteristic of barbel chub, *Squaliobarbus curriculus* provides insights into antiviral immunity against infection with grass carp reovirus. *Fish & Shellfish Immunology*, 66, 120-128.
- Watanuki, H., Gushiken, Y., Takahashi, A., Yasuda, A., & Sakai, M. (2000). *In vitro* modulation of fish phagocytic cells by β -endorphin. *Fish & Shellfish Immunology*, 10(2), 203-212.
- Watkins, R. R., & Bonomo, R. A. (2016). Overview: Global and local impact of antibiotic resistance. *Infectious Disease Clinics*, 30(2), 313-322.
- Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Geer, L. Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Ostell, J., Miller, V., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R. L., Tatusova, T. A., Wagner, L., & Yaschenko, E. (2007). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 36(suppl_1), D13-D21.
- Whyte, S. K. (2007). The innate immune response of finfish—a review of current knowledge. *Fish & Shellfish Immunology*, 23(6), 1127-1151.
- Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics*, 3(10), e000132.

- Wiklund, T., & Dalsgaard, I. (1998). Occurrence and significance of atypical *Aeromonas salmonicida* in non-salmonid and salmonid fish species: A review. *Diseases of Aquatic Organisms*, 32(1), 49-69.
- Willett, C. E., Cortes, A., Zuasti, A., & Zapata, A. G. (1999). Early hematopoiesis and developing lymphoid organs in the zebrafish. *Developmental Dynamics: An Official Publication of the American Association of Anatomists*, 214(4), 323-336.
- Williams, P., Winzer, K., Chan, W. C., & Camara, M. (2007). Look who's talking: Communication and quorum sensing in the bacterial world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1483), 1119-1134.
- Willis, D. K., Hrabak, E. M., Rich, J. J., Barta, T. M., Lindow, S. E., & Panopoulos, N. J. (1990). Isolation and characterization of a *Pseudomonas syringae* pv. *syringae* mutant deficient in lesion formation on bean. *Molecular Plant-Microbe Interactions*, 3(3), 149-156.
- Wolska, K., Jakubczak, A., Anusz, Z., & Bukowski, K. (1999). Susceptibility of *Pseudomonas aeruginosa* strains to antibiotics and chemotherapy. *Medycyna Weterynaryjna*, 55(12), 812-817.
- World Bank. (2013). Fish to 2030: Prospects for fisheries and aquaculture (English). Retrieved from <http://documents.worldbank.org/curated/en/45863146815237668/pdf/831770WP0P11260ES003000Fish0to02030.pdf>
- World Health Organization. (2020). *Global antimicrobial resistance surveillance system (GLASS) report: Early implementation 2020*. World Health Organization.

- Xie, Z. W., Xu, W. F., Davie, E. W., & Chung, D. W. (1998). Molecular cloning of human ABPL, an actin-binding protein homologue. *Biochemical and Biophysical Research Communications*, 251(3), 914-919.
- Xu, P., Zhang, X., Wang, X., Li, J., Liu, G., Kuang, Y., Xu, J., Zheng, X., Ren, L., Wang, G., Zhang, Y., Huo, L., Zhao, Z., Yu, D., Xu, L., Wang, B., Li, Y., Zhu, Y., Xue, W., Zhao, L., Wang, J., Gu, Y., Lv, W., Wu, K., Xiao, J., Wu, J., Zhang Z., Yu, J., & Sun, X. (2014). Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nature Genetics*, 46(11), 1212-1219.
- Xu, W. T., & Chen, S. L. (2017). Genomics and genetic breeding in aquatic animals: progress and prospects. *Frontiers of Agricultural Science and Engineering*, 4(3), 305-318.
- Yang, Y., Kitagaki, J., Wang, H., Hou, D. X., & Perantoni, A. O. (2009). Targeting the ubiquitin-proteasome system for cancer therapy. *Cancer Science*, 100(1), 24-28.
- Yano, T. (1996). The nonspecific immune system: Humoral defense. *The Fish Immune System: Organism, Pathogen, and Environment*, 105-157.
- Yao, C. L., Kong, P., Wang, Z. Y., Ji, P. F., Cai, M. Y., Liu, X. D., & Han, X. Z. (2008). Cloning and expression analysis of two alternative splicing toll-like receptor 9 isoforms A and B in large yellow croaker, *Pseudosciaena crocea*. *Fish & Shellfish Immunology*, 25(5), 648-656.
- Ye, S., Dhillon, S., Ke, X., Collins, A. R., & Day, I. N. (2001). An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Research*, 29(17), e88-e88.

- Yuan, Z., Liu, S., Zhou, T., Tian, C., Bao, L., Dunham, R., & Liu, Z. (2018). Comparative genome analysis of 52 fish species suggests differential associations of repetitive elements with their living aquatic environments. *BMC Genomics*, 19(1), 1-10.
- Yue, G. H. (2014). Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries*, 15(3), 376-396.
- Yue, G. H., & Wang, L. (2017). Current status of genome sequencing and its applications in aquaculture. *Aquaculture*, 468, 337-347.
- Yue, X., Wang, H., Huang, X., Wang, C., Chai, X., Wang, C., & Liu, B. (2012). Single nucleotide 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 Immunology*, 33(3), 559-568.
- Zaman, S. B., Hussain, M. A., Nye, R., Mehta, V., Mamun, K. T., & Hossain, N. (2017). A review on antibiotic resistance: Alarm bells are ringing. *Cureus*, 9(6).
- Zapata, A. G., Varas, A., & Torroba, M. (1992). Seasonal variations in the immune system of lower vertebrates. *Immunology Today*, 13(4), 142-147.
- Zapata, A., Diez, B., Cejalvo, T., Gutierrez-de Frias, C., & Cortés, A. (2006). Ontogeny of the immune system of fish. *Fish & Shellfish Immunology*, 20(2), 126-13.
- 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. *Frontiers in Genetics*, 10, 700.

- Zhao, R., Feng, J., Liu, J., Fu, W., Li, X., & Li, B. (2019). Deciphering of microbial community and antibiotic resistance genes in activated sludge reactors under high selective pressure of different antibiotics. *Water Research*, 151, 388-402.
- 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.
- Zhenzhen, X., Ling, X., Dengdong, W., Chao, F., Qiongyu, L., Zihao, L., Xiaochun, L., Yong, Z., Shuisheng, L., & Haoran, L. (2014). Transcriptome analysis of the *Trachinotus ovatus*: Identification of reproduction, growth and immune-related genes and microsatellite markers. *PLOS One*, 9(10), e109419.
- 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), 362-369. Thieme Medical Publishers.
- Zuko, A., Kleijer, K. T., Oguro-Ando, A., Kas, M. J., van Daalen, E., van der Zwaag, B., & Burbach, J. P. H. (2013). Contactins in the neurobiology of autism. *European Journal of Pharmacology*, 719(1-3), 63-74.

APPENDICES



Figure S1. Healthy sample of Empurau isolated from the LTT Aquaculture Sdn Bhd.

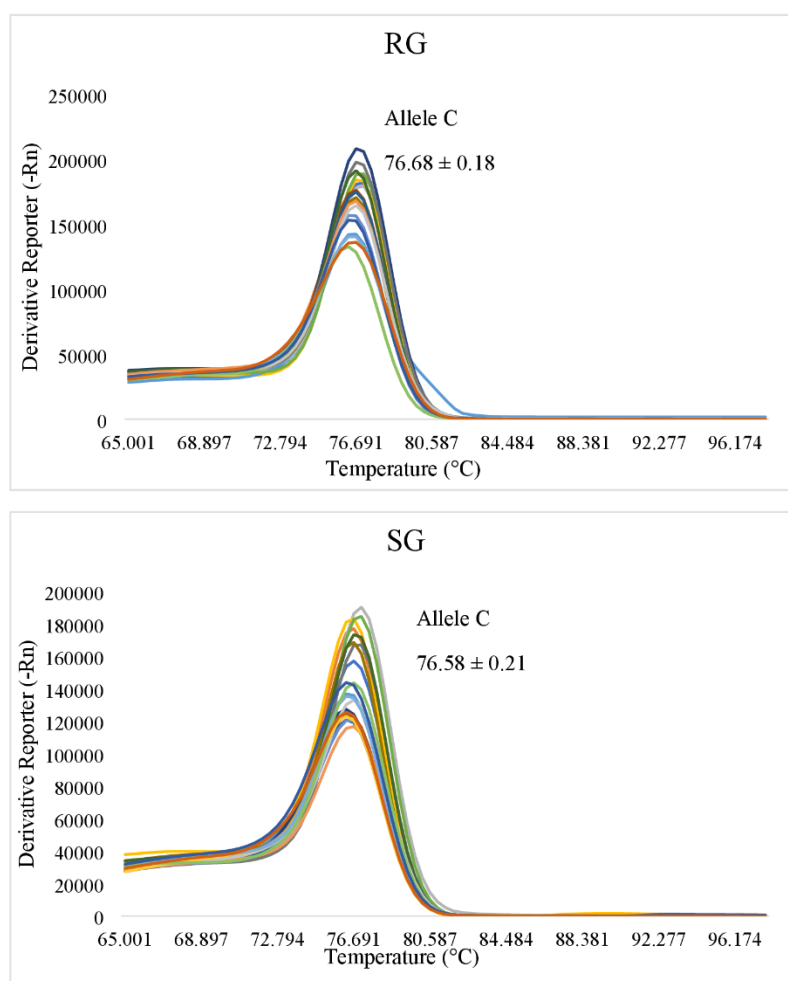


Figure S2. It showed the melt curve for all the twenty samples from (A) RG and (B) SG in 6151 C/T SNP. Allele C had shown T_m of 76.68 ± 0.18 (RG) and 76.58 ± 0.21 (SG)

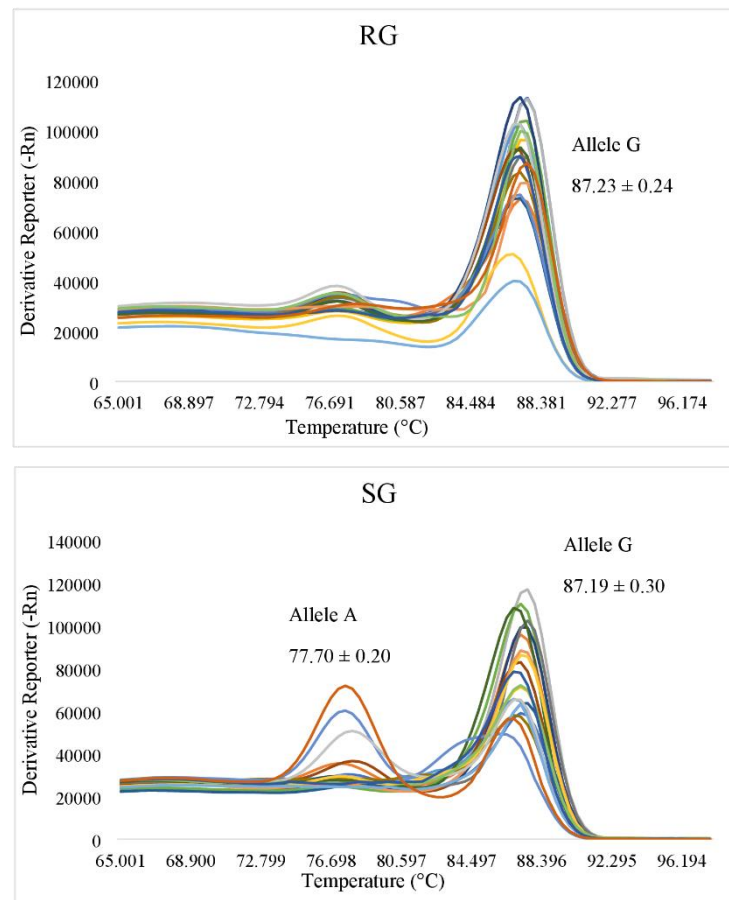


Figure S3. It showed the melt curve for all the twenty samples from (A) RG and (B) SG in 112388 G/A SNP. Allele A had shown T_m of 87.23 ± 0.24 (RG) and 77.70 ± 0.20 (SG) while allele G had shown T_m of 87.19 ± 0.30

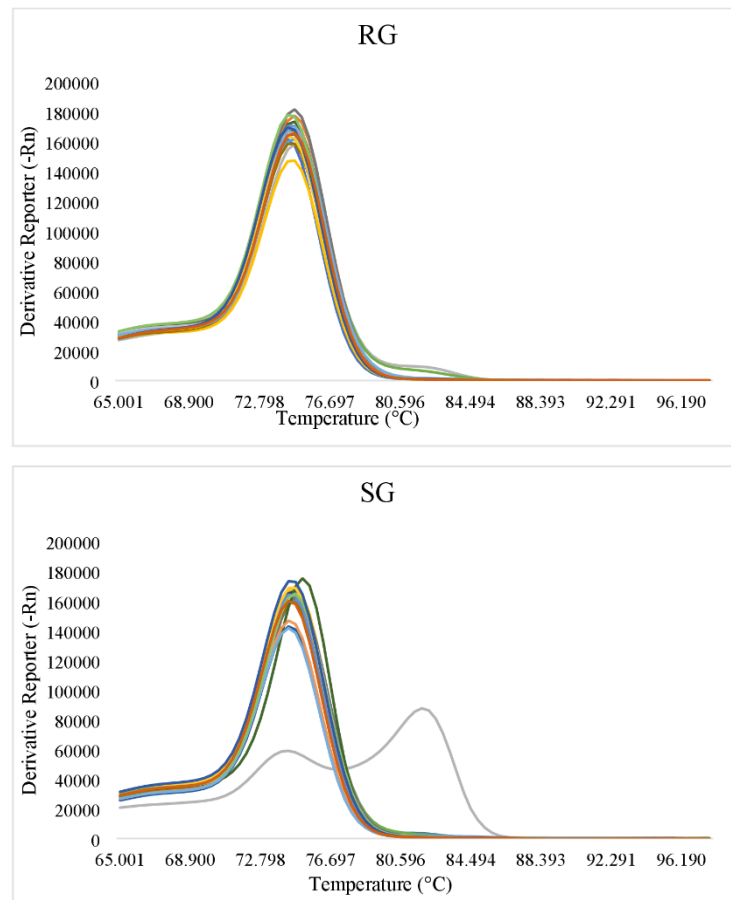


Figure S4. It showed the melt curve for all the twenty samples from (A) RG and (B) SG in 59747 G/A SNP. RG showed a single peak at T_m of 74.59 ± 0.12 while most of the samples of SG showed a single peak at 74.59 ± 0.18 with an exception of a single sample showing double peak at the T_m of 81.74

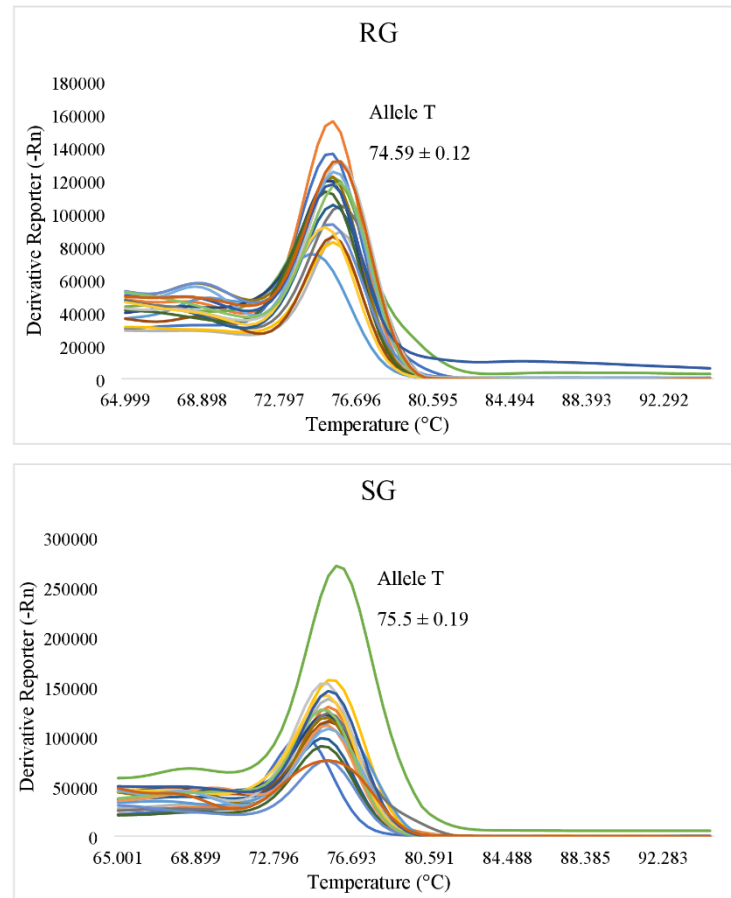


Figure S5. It showed the melt curve for all the twenty samples from (A) RG and (B) SG in 27857 A/T SNP. Allele T had shown T_m of 74.59 ± 0.12 (RG) and 75.5 ± 0.19 (SG)

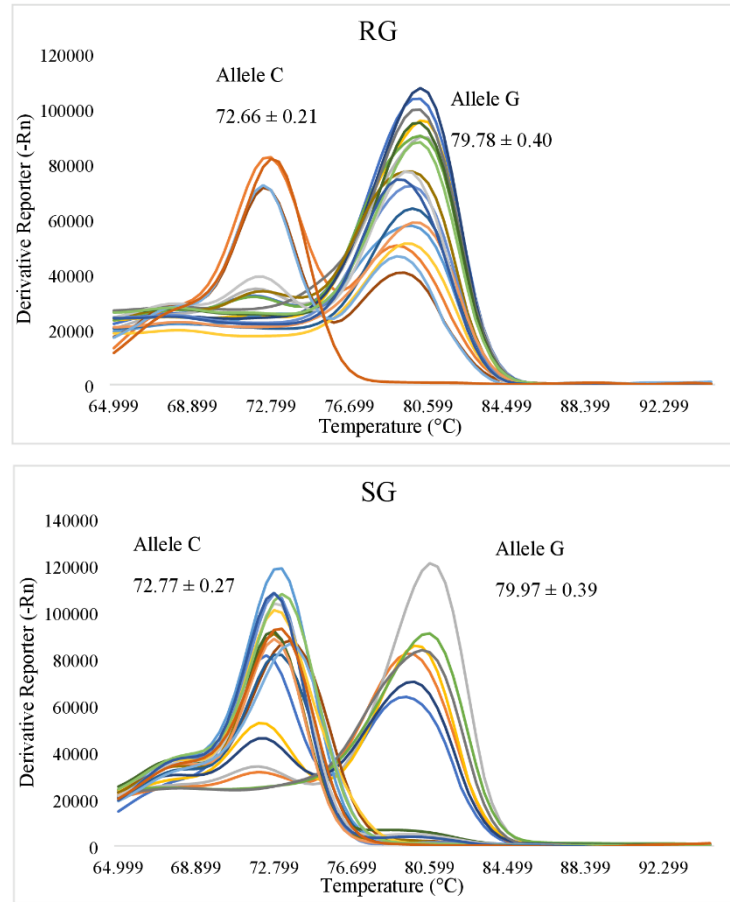


Figure S6. It showed the melt curve for all the twenty samples from (A) RG and (B) SG in 19564 G/C SNP. Allele G had shown T_m of 79.78 ± 0.40 (RG) and 79.97 ± 0.39 (SG) while allele C shown T_m of 72.66 ± 0.21 (RG) and 72.77 ± 0.27 (SG)

