





Article

Development of Disease-Resistance-Associated Microsatellite DNA Markers for Selective Breeding of Tilapia (*Oreochromis* spp.) Farmed in Taiwan

Che-Chun Chen^{1,2}, Chang-Wen Huang^{1,3,4} , Chung-Yen Lin^{2,5,6} , Chia-Hui Ho³, Hong Nhat Pham^{3,7}, Te-Hua Hsu^{3,4} , Tzu-Tang Lin⁵, Rong-Hwa Chen⁸, Shuenn-Der Yang⁸, Chin-I. Chang⁹ and Hong-Yi Gong^{1,3,4,*} 

- ¹ Doctoral Degree Program in Marine Biotechnology, National Taiwan Ocean University, Keelung 20224, Taiwan; fassadze@gmail.com (C.-C.C.); cw Huang@mail.ntou.edu.tw (C.-W.H.)
- ² Doctoral Degree Program in Marine Biotechnology, Academia Sinica, Taipei 11529, Taiwan; cylin@iis.sinica.edu.tw
- ³ Department of Aquaculture, National Taiwan Ocean University, Keelung 20224, Taiwan; gyyuty1@yahoo.com.tw (C.-H.H.); hongnhat@ria1.org (H.N.P.); realgigi@mail.ntou.edu.tw (T.-H.H.)
- ⁴ Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung 20224, Taiwan
- ⁵ Institute of Information Science, Academia Sinica, Taipei 11529, Taiwan; tzutang@iis.sinica.edu.tw
- ⁶ Institute of Fisheries Science, National Taiwan University, Taipei 10617, Taiwan
- ⁷ Research Institute for Aquaculture No. 1 (RIA1), Dinh Bang, Tu Son 16352, Bac Ninh, Vietnam
- ⁸ Freshwater Aquaculture Research Center, Fisheries Research Institute, Council of Agriculture, Changhua 50562, Taiwan; rhchen@mail.tfrin.gov.tw (R.-H.C.); sdyang@mail.tfrin.gov.tw (S.-D.Y.)
- ⁹ Fisheries Research Institute, Council of Agriculture, Keelung 20246, Taiwan; cichang@mail.tfrin.gov.tw
- * Correspondence: hygong@mail.ntou.edu.tw



Citation: Chen, C.-C.; Huang, C.-W.; Lin, C.-Y.; Ho, C.-H.; Pham, H.N.; Hsu, T.-H.; Lin, T.-T.; Chen, R.-H.; Yang, S.-D.; Chang, C.-I.; et al. Development of Disease-Resistance-Associated Microsatellite DNA Markers for Selective Breeding of Tilapia (*Oreochromis* spp.) Farmed in Taiwan. *Genes* **2022**, *13*, 99. <https://doi.org/10.3390/genes13010099>

Academic Editor: Antonio Figueras

Received: 16 December 2021

Accepted: 29 December 2021

Published: 31 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: There are numerous means to improve the tilapia aquaculture industry, and one is to develop disease resistance through selective breeding using molecular markers. In this study, 11 disease-resistance-associated microsatellite markers including 3 markers linked to *hamp2*, 4 linked to *hamp1*, 1 linked to *pgrn2*, 2 linked to *pgrn1*, and 1 linked to *piscidin 4* (*TP4*) genes were established for tilapia strains farmed in Taiwan after challenge with *Streptococcus iniae*. The correlation analysis of genotypes and survival revealed a total of 55 genotypes related to survival by the chi-square and Z-test. Although fewer markers were found in B and N2 strains compared with A strain, they performed well in terms of disease resistance. It suggested that this may be due to the low potency of some genotypes and the combinatorial arrangement between them. Therefore, a predictive model was built by the genotypes of the parental generation and the mortality rate of different combinations was calculated. The results show the same trend of predicted mortality in the offspring of three new disease-resistant strains as in the challenge experiment. The present findings is a nonkilling method without requiring the selection by challenge with bacteria or viruses and might increase the possibility of utilization of selective breeding using SSR markers in farms.

Keywords: tilapia; disease resistance; *Streptococcus iniae*; gene amplification; hepcidin; progranulin; piscidin; microsatellite; marker-assisted selection; predictive model

1. Introduction

Tilapia comprises over 100 species of freshwater fish in Africa that belong to the genera *Oreochromis*, *Sarotherodon*, and *Tilapia*. Seventy-five countries have economies of scale for feeding Nile tilapia. Tilapia is one of the most important economic farmed fish in Taiwan, as well as in the world. According to the Food and Agriculture Organization (FAO), tilapia is one of the important species for providing protein in terms of the human diet [1]. Selective breeding programs have been widely used in aquatic farming (more than 60 species) for fish such as common carp, grass carp, rainbow trout, Atlantic salmon, channel catfish,

sea bream, oyster, shrimp, and tilapia [2]. The GIFT strain (Genetic Improvement of Farmed Tilapia) in Nile tilapia (*Oreochromis niloticus*) started in 1989 and originated from the GIFT selective breeding project in the Philippines [3]. While the GIFT strain is well known for its superior growth performance, it is characterized by a comparatively weakened immune system [4].

Diseases associated with Nile tilapia have gathered serious attention in recent years, including viral encephalitis of tilapia larvae (primary characterization of a novel herpes-like virus) and TLEV (tilapia larva encephalitis virus) [5,6]. Many Gram-positive and Gram-negative bacteria can infect tilapia, such as of the following genera: *Aeromonas*, *Citrobacter*, *Edwardsiella*, *Flavobacterium*, *Pseudomonas*, *Streptococcus*, and *Mycobacterium* [6,7]. Among these, *Streptococcus agalactiae* and *Streptococcus iniae* are the major bacterial pathogens in tilapia [8]. To solve problems of the related diseases, research has sought to improve the immune capacity of fish through the upregulation expression of antimicrobial peptides (AMPs) [9]. AMP expression is one mechanism used by the innate immune system of fish [10] that can effectively inhibit the growth of bacteria, and even improved survival after the injection of synthetic AMPs has been observed [11]. In fish, most AMPs are α -helical peptides [12]. Moreover, a total of 122 fish AMPs have been reported, including cathelicidin, chrysopsin, dicentracin, epinecidine, hepcidin, misgurin, oncorhynsin, and piscidin [13–16]. In contrast to blue tilapia (*Oreochromis aureus*), Nile tilapia has stronger disease resistance [17]. Another study also revealed that Nile tilapia had more AMP genes than blue tilapia (*Oreochromis aureus*), especially hepcidin [18]. Additionally, gene diversification and amplification processes have occurred in fish hepcidin [19]. These results suggest a correlation between disease resistance and AMPs.

Hepcidin is a regulatory factor of iron and also a hepatic antimicrobial peptide (HAMP) that is primarily expressed in the liver [20,21]. It has been extensively studied in various species, including fish. Three hepcidins have been found in *Oreochromis mossambicus*, namely Th1-5, Th2-2, and Th2-3 [21]. Two AMPs, Hep-JF1 and Hep-JF2, were found in Japanese flounder (*Paralichthys olivaceus*). Six hepcidin genes (LcHamps) with diversified regulation and functions involved in antibacterial activity, antiviral activity, and regulation of intracellular iron metabolism were identified in large yellow croaker (*Larimichthys crocea*) genome [19]. Hepcidin showed the strongest expression in the liver after LPS induction on the convict cichlid (*Archocentrus nigrofasciatus*) [11]. Studies of pathogenic bacteria treated with synthetic hepcidin peptides showed that the pathogen cell membranes were destroyed. Studies also indicated that oral intake or injection of hepcidin in zebrafish could inhibit bacterial growth and stimulate the host immune response, which significantly improved their survival rate following pathogen infection [22,23]. Hepcidin was used against *Staphylococcus aureus* and *Vibrio vulnificus* in spotted grouper (*Epinephelus coioides*) [24]. Pagaporn et al. [25] mentioned that hepcidin could inhibit both virulence and growth of bacteria by reducing the level of metals (e.g., Fe) in fish tissues. The study indicated that hepcidin enhanced streptococcosis resistance in Nile tilapia.

Simple sequence repeats (SSRs), also known as microsatellites or short tandem repeats (STRs), are tandem repeats of short sequence motifs commonly found in eukaryotic genomes. They are also considered as useful genetic markers for genetic diversity analysis, DNA fingerprinting, and linkage mapping. Several studies have shown that the distribution of SSRs in the genome is non-random. There are more SSRs in the untranslated regions (UTRs) than in the coding regions and are thought to be associated with the regulation of gene expression [26]. Recent studies have revealed that microsatellite polymorphisms on a genome-wide scale contribute to the heritability of human gene expression. Evidence has been found for the role of microsatellites in regulating transcription factor binding, methylation, promoter, enhancers, mRNA stability, alternative splicing, nucleosome modification, and noncoding RNA [27].

With the development of genetic technology, marker-assisted selection has been applied in many aquaculture species [28]. However, research developing microsatellite markers of disease resistance for fish is scarce. In this study, disease-resistance-associated

microsatellite markers were developed from the Nile tilapia genome. To confirm that the microsatellite markers were associated with disease resistance, seven populations were chosen and analyzed by *Streptococcus* infection. As far as we know, this is the first report describing the development of disease-resistance-associated microsatellites in Taiwan tilapia. Thus, we hope our findings will have a positive practical impact on the tilapia industry in Taiwan.

2. Materials and Methods

2.1. Experimental Animals

Seven tilapia strains were used in this study. The NT1 strain is a Nile tilapia (*Oreochromis niloticus*) strain no.1 of National Taiwan Ocean University [29]. Two commercial Taiwan tilapia (*Oreochromis* spp.) strains, A and B strains, were collected from fish farms in south Taiwan. The Freshwater Aquaculture Research Center, Fisheries Research Institute provided the Nile tilapia (*Oreochromis niloticus*) N2 strain all-male XY fish obtained from crossing of YY supermale and XX female derived from N2 strain [30]. ANT1 (A male × NT1 female), AB (A male × B female) hybrid strains, and BB (B male × B female) strain are new *Streptococcus*-resistant strains. The transcriptome was generated through the NT1 strain and used for the selection of disease-resistance-related genes and SSR markers. SSR markers verification and the disease-resistant parental generation was performed by A, B, and N2 strains. Genetic inheritance of disease-resistant genotypes and confidence of predictive models were examined by ANT1, AB, and BB strains (Figure 1). All fish strains were cultured in a tank in 2T water at a stable temperature of 28 °C and were fed twice a day.

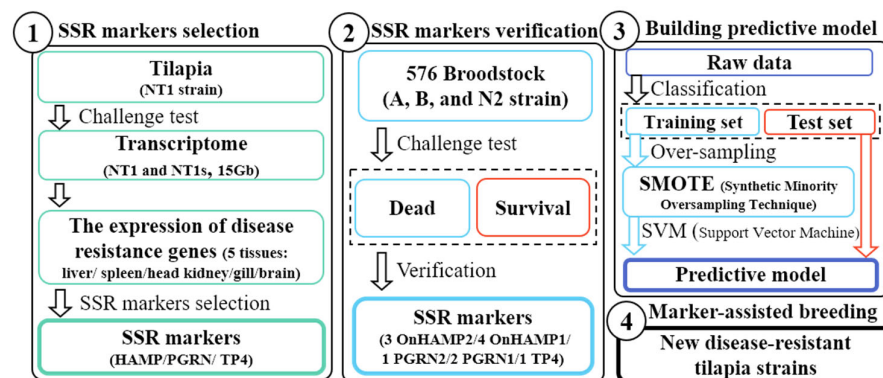


Figure 1. Experimental scheme showing the four main axes in this study. (1) SSR markers selection: transcriptome data were created by the NT1 strain challenge test, and SSR markers were selected through the transcriptome and gene expression results (NT1: NT1 strain without infection; NT1S: NT1 strain infected by 10^4 CFU/g *Streptococcus iniae*). (2) SSR markers verification: a total of 576 fish from A, B, and N2 strains were infected with *S. iniae* to verify the association of SSR markers with disease resistance. (3) Building the predictive model: the raw data from step 2 were over-sampled by SMOTE method (synthetic minority oversampling technique), and then used to establish a predictive model through SVM (support vector machine). (4) Marker-assisted breeding: establish new disease-resistant tilapia strain by the marker-assisted breeding method.

2.2. Transcriptome Generation

2.2.1. Sample Collection

To generate tilapia transcriptome, the Nile tilapia NT1 strains (10 g, three months old) were prepared. NT1: NT1 strain without infection; NT1S: NT1 strain infected by *Streptococcus iniae* 89353 (10^4 CFU/g body weight) [31] at 12 hpi. At 12 hpi by *S. iniae* 89353, five 3-month-old whole fish (10 g) from each group were sacrificed for total RNA extraction.

2.2.2. Extraction of Total RNA

Approximately 0.1 g each of the liver, spleen, head kidney, gill, and brain tissues were individually collected in a 1.5 mL centrifuge tube. We then added 1 mL TRIzol and homogenized the mixture. The samples were incubated at room temperature for 5 min and centrifuged at $10,000\times g$ for 2 min at 4 °C. Following this, we added 0.2 mL chloroform followed by vigorous shaking for 15 s and incubation at room temperature for 3 min. We then centrifuged the mixture at $10,000\times g$ for 15 min at 4 °C. The supernatant liquid was transferred to a new centrifuge tube, 0.5 mL isopropanol was added, and the tube was mixed through gentle shaking. After incubation at room temperature for 10 min, the mixture was centrifuged at $10,000\times g$ for 15 min at 4 °C, and the supernatant liquid was discarded. We then added 400 μ L of 70% alcohol to the tube, following by gentle shaking and then centrifugation at $7500\times g$ for 10 min at 4 °C. The supernatant was removed, and the sample was placed at 55 °C in an incubator for 3 min. Next, 50 μ L DEPC-treated ddH₂O (diethylpyrocarbonate-treated ddH₂O) was added and the sample incubated at 55 °C for 15 min. A spectrophotometer was used to measure the ratio of OD₂₆₀ to OD₂₈₀ (with 1.9–2.0 corresponding to high purity). Finally, the samples were stored at –80 °C for later use.

2.2.3. Purification of Total RNA

Using the PureLink RNA Mini Kit for purification of total RNA, we added 600 μ L lysis buffer to the total RNA sample and vortexed until the mixture was homogenous. Then, we added 600 μ L of 70% alcohol and vortexed evenly. A maximum of 600 μ L of the mixture was transferred into a spin column, centrifuged at $12,000\times g$ for 15 min at 4 °C, and this procedure repeated until the entire mixture was processed. Next, 350 μ L Wash Buffer I was added, followed by centrifugation at $12,000\times g$ for 25 s at 4 °C, and the filtrate was removed. Following this, 80 μ L DNase I mixture (10 μ L DNase and 70 μ L RDD Buffer) was added followed by incubation at room temperature for 15 min. Then, another 350 μ L Wash Buffer I was added, followed by centrifugation $12,000\times g$ for 25 s at 4 °C, and the filtrate was removed. Next, 500 μ L Wash Buffer II was centrifuged (at $12,000\times g$ for 25 s at 4 °C, and the filtrate was removed by centrifuging at $12,000\times g$ for 5 min at 4 °C; this procedure was repeated. The spin column was transferred into a recovery tube, and 20–50 μ L of RNase-free water was then added, followed by incubation at room temperature for 2 min and centrifugation at $12,000\times g$ for 2 min at 4 °C. RNA detection and quantification was conducted using a NanodropTM 1000 spectrophotometer.

2.2.4. Assembly, Function Annotation, and Differential Gene Expression

Trinity 12 software (r20140717, Broad Institute) [32] was used for de novo genome-guided assembly. The sequence used for reference was the published draft genome sequence of Nile tilapia/oreNil2 (oreNil2, Broad Institute of MIT and Harvard Orenil1.1 (GCA_000188235.1)). De novo transcriptome assembly of NT1 was submitted to the NCBI short read archive database (accession numbers: SRR14141863 and SRR14141864).

The BAM file was generated from read mapping to the genome by TopHat (v2.0.13) [33] and the BAM file was assembled with Trinity 12. Subsequently, we mapped the reads to contigs using bowtie2 software (v2.2.3) [34]. The standardized count and FPKM of expression for all transcripts were calculated using RSEM (v1.2.0) [35]. We then calculated the expression and filtered the contigs so that the FPKM value was greater than 0.01. These contigs were compared with the Nile tilapia database (Oreochromis niloticus.Orenil1.0.cdna.all.fa from http://ftp.ensembl.org/pub/release-78/fasta/oreochromis_niloticus/cdna/ accessed on 7 May 2015) through BLASTn (NT) ($p < 0.00001$).

The transcript DNA sequence was analyzed or compared using the vertebrate mammalian (VBMM: ftp.ncbi.nlm.nih.gov/refseq/release/vertebrate_mammalian/), the vertebrate non-mammalian (VBnonMM: http://ftp.ncbi.nlm.nih.gov/refseq/release/vertebrate_other/ accessed on 7 May 2015), SwissProt, KOG (EuKaryotic Orthologous Groups), KEGG (Kyoto Encyclopedia of Genes and Genomes), and GO (Gene Ontology) databases. The GO,

KOG, and KEGG categories were annotated with BLASTX (with criteria: e value $< 10^{-5}$) and the Blast2GO [36]. Finally, analysis of differential expression between any two samples was conducted by edgeR (V 3.10.3) [37].

2.3. The Gene Expression of Hamp in Tilapia

2.3.1. Reverse Transcription

Total RNA (extracted and purified as described in Sections 2.2.2 and 2.2.3) was reverse transcribed using a high-capacity cDNA reverse transcription kit. The sample was diluted to 100 ng/ μ L. We then mixed 10 μ L RNA sample, 2 μ L 10 \times RT Buffer, 0.8 μ L 25 \times dNTP mix (100 mM), 2 μ L 10X RT Random primer, 1 μ L MultiScribe™ Reverse Transcriptase (50 U/ μ L), and 4.2 μ L DEPC-treated ddH₂O. PCR was conducted using a thermocycler (TProfessional Thermocycler, Biometra, BM-070-801), and the reaction conditions were the following: 1. 25 °C for 10 min; 2. 37 °C for 120 min; 3. 85 °C for 5 min. The sample was diluted 50X and stored at –20 °C for later use.

2.3.2. Real-Time Quantitative PCR, qPCR

We used a Fast SYBR® Green PCR Master for qPCR. The total volume was 20 μ L and included 10 μ L 2 \times Fast SYBR® Green PCR Master Mix, 1 μ L of 4 μ M forward primer, 1 μ L of 4 μ M reverse primer, 3 μ L DEPC-treated ddH₂O, and 5 μ L cDNA. The reaction included four steps: 1. 50 °C for 2 min; 2. 95 °C for 10 min; 3. 95 °C for 15 s and 60 °C for 60 s, repeated 40 times; 4. 95 °C for 15 s, 60 °C for 60 s, 95 °C for 15 s, and 60 °C for 15 s.

2.4. Selecting the Disease-Resistance-Associated Microsatellites

We searched for disease-resistance-associated microsatellites in the Nile tilapia genome (NCBI). Microsatellites can be classified on the basis of their size of the repeated sequence as follows: single nucleotide microsatellites ≥ 10 bp, dinucleotide microsatellites ≥ 6 bp, trinucleotide microsatellites ≥ 5 bp, tetranucleotide microsatellites ≥ 5 bp, pentanucleotide microsatellites ≥ 5 bp, and hexanucleotide microsatellites ≥ 5 bp [38]. We designed a microsatellite-specific primers using an online tool (Websat, <http://wsmartins.net/websat/> accessed on 7 May 2015); primer length was 22 bp, T_m was 60 °C, GC was 60%, and the product length was between 100 and 400 bp [39].

2.5. Streptococcus iniae Challenge

2.5.1. Collecting and Weighing Samples

Nile tilapia (body weight: 10–80 g) were used for the challenge experiment. The experimental group included F₀ (NT1, A, B, and N2) and F₁ (ANT1, AB, and BB) groups. To assess the effect of disease resistance, 20 fish were injected with PBS control at the same time. The fish were maintained in 30 L tanks at 28 °C following pathogen challenge. We continued the observation until no dead fish were observed for at least for three days. The fish were then used in the following experiments.

2.5.2. Streptococcus iniae Culture

Liquid BHI was prepared for *Streptococcus iniae* 89353 cultures. A total of 30 μ L *Streptococcus iniae* 89353 and 3 mL BHI broth were added into the autoclaved conical container. The mix liquid culture was left to grow overnight at 30 °C for 16 h. After 16 h, we cultured the entire aliquot of *S. iniae* in a larger volume for 8 h. This exact volume depended on the fish weight and the applied dose in subsequent procedures. The *S. iniae* 89353 bacteria were provided by the Bureau of Animal and Plant Health Inspection and Quarantine, Council of Agriculture, Executive Yuan, Kaohsiung Branch Dr. Benjia Zhao. Bacteria were re-isolated from a single symptomatic fish for biochemical confirmation of their identity.

2.5.3. Intraperitoneal Injection

After anesthetizing the fish with 1 mL 2-phenoxyethanol ($\geq 99\%$), the *S. iniae* was intraperitoneally injected into the tilapia with a lethal dose (LD) of 70–80 (strain A is 2×10^6 ; B is 6×10^5 ; N2 is 6.5×10^5 CFU/g body weight). The control group was injected with PBS at the same volume of *S. iniae*. *S. iniae* was cultured in liquid BHI overnight (*Streptococcus iniae* culture as described in Section 2.5.2). Next, the dose was accurately calculated. The fish were monitored every day for the occurrence of death, and in this case, about 0.1 g of caudal fin was cut with sterilized scissors and placed into 70% alcohol for preservation. We continued observation until no dead fish were observed for at least three days. Finally, we cut approximately 0.1 g of fins from the surviving fish and stored samples in 70% alcohol.

2.6. Genomic DNA Extraction

Whole-genome DNA was extracted by a MasterPure™ DNA Purification Kit. First, we removed the alcohol from the fin samples of tilapia after challenge and placed the sample into a new tube. We then added 300 μ L tissue and cell lysis solution containing 50 μ g of proteinase K, followed by thorough mixing and incubation of the solution at 55 °C for 16 h. Using a blunt pipette tip, we transferred the supernatant to a fresh microcentrifuge tube, added 175 μ L of MPC protein precipitation reagent, and mixed the solution gently, avoiding vortexing, as it may cause genomic DNA breakage. We centrifuged this mixture at $5000 \times g$ for 10 min at room temperature. We again used a blunt pipette tip to transfer the supernatant to a new microcentrifuge tube. Then, we added 500 μ L of 100% isopropanol followed by thorough mixing. White filiform genomic DNA was visible during mixing, which was picked up using the tip and placed into a new microcentrifuge tube, and 500 μ L of 70% ethanol was then added, followed by incubation at room temperature for 5 min. Ethanol was then removed, and the tube containing the DNA was placed in an incubator at 50 °C until the genomic DNA pellet turned transparent. Following this, 200–500 μ L TE buffer was added and this mixture was incubated at 50 °C; gel electrophoresis was then used to confirm the quality of genomic DNA. The DNA concentration was measured using the nanodrop and the concentration of the solution was adjusted to 25 ng/ μ L before storage at -20 °C.

2.7. PCR Amplification of Fluorescently Labeled Microsatellite Products Using Specific Primers

The method to prepare fluorescently labeled microsatellite product was modified from standard multiplex PCR [40]. Primer details are shown in Table 1. First, we added 2 μ L of 25 ng/ μ L DNA, 2.5 μ L of $10 \times$ PCR-MgCl₂ buffer, 0.5 μ L of a 10 mM dNTP mixture, 0.75 μ L of 50 mM MgCl₂, 0.5 μ L Platinum® Taq polymerase, 0.5 μ L of 10 μ M forward primer, 0.5 μ L of 10 μ M reverse primer, and 17.75 μ L ddH₂O to the reaction tube (the total volume was 25 μ L). The mixture was briefly centrifuged before PCR using a thermocycler (TProfessional Thermocycler, Biometra, BM-070-801). The reaction conditions were as follows: 1. 95 °C for 5 min; 2. 95 °C for 30 s; 3. 60 °C for 30 s; 4. 72 °C for 1 min; 5. 72 °C for 10 min. Next, steps 2–4 were repeated 35 times. This PCR reaction adds a G/C-rich adaptor sequence of 17 bases (5'GAGCACGAGGAGA3') via inclusion in the 5' site of all primers and serves as a site for binding site in the second PCR. Next, we took 2 μ L of amplified product as a template and a fluorescently labeled primer was used as the forward primer for 5' labeling of the product produced by the second polymerase chain reaction. The reaction time of the 3' primer (reaction drug) was same as the first PCR reaction. The selection of products that were successfully labeled with the fluorescent primer was based on product size; products labeled with different fluorescent tags were similarly sized. The final product was stored at -20 °C.

2.8. Sequencing and Genotyping

We then checked the expected size of the product after microsatellite marker analysis. Next, the Digimage system (Major Science Digimage System DI-HD-110) was used to

record images. To avoid excessive differences in fluorescence signals, we first used a Synergy™ Multi-detection microplate reader to measure the concentration of each sample. Second, we quantified the concentration of each sample and adjusted samples to the same concentration. Third, we carried out equal volume mixing. The mixing was based on the same fluorescent marker of microsatellite product size, with differences of about 100–150 bp. Finally, we mixed the plural samples for the basic fluorescent peak and fluorescence signal analysis with an ABI 3730XL DNA analyzer. After the analysis, we used Genemarker software to analyze the microsatellite products size with different fluorescent markers and determined the genotype of individual fish.

2.9. Calculations and Statistical Analysis

2.9.1. The Genetic Diversity Analysis

Genetic diversity analysis of tilapia populations with microsatellite markers was conducted using computer software GenAlEx 6.5 and the online tool Gene Calc (<https://www.gene-calc.pl/pic>). We included the number of alleles (N_a), the effective number of alleles (N_e), Shannon's information index (I), allele frequency, genotype frequency, observed heterozygosity (H_o), expected heterozygosity (H_e), unbiased expected heterozygosity (uH_e), polymorphic information content (PIC), individual fixation indices (F_{IS} , F_{IT} , and F_{ST}), and the number of migrants per generation (N_m) [41].

2.9.2. The Correlation Analysis of Genotype and Survival Rate

The different genotypes were determined in the different groups, and the death or survival of fish after challenge was taken as a variable. Then, we performed Pearson's chi-square test, correlation analysis, and the Z-test using IBM SPSS Statistics version 25.0 software (SPSS Inc., Chicago, IL, USA). Moreover, post hoc Bonferroni correction was used to adjust the p value. $\alpha 0k = \alpha 0'$ was mentioned in the Bonferroni correction, where k represents the total examination number.

2.10. The Effectiveness Analysis of Genotypes by Predictive Models

This technique uses a support vector machine [42] to build the predictive model with an IBM SPSS modeler subscription (version SaaS). The process of building and applying a predictive model has three basic steps: 1. Building a predictive model; 2. Testing the predictive model; 3. Applying the predictive model.

2.10.1. Building a Predictive Model

The predictive model established in this study predicts the number of death events following *S. iniae* challenge, and also predicts the mortality rate. A total of 384 fish from the A and B strains were selected to train the model. Eight predictors were chosen, including SSR2, SSR4, SSR7, SSR14, SSR18, SSR19, SSR21, and SSR22. The event of death was chosen as the target (survival: 0; death: 1). In total, 90% of all samples were randomly selected as the training set and 10% were chosen as the testing set. To balance the efficacy of both the dead and alive groups, the synthetic minority oversampling technique [43] was used as an imbalanced data processing method. The testing set was an independent set and was not used for SMOTE and training purposes.

2.10.2. Evaluation of the Predictive Model

First, we made a confusion matrix containing a true positive (TP), a true negative (TN), a false positive (FP), and a false negative (FN). These four parameters were then used to calculate the true positive rate (TPR), the true negative rate (TNR), the positive predictive value (PPV), the negative predictive value (NPV), the false positive rate (FPR), the false discovery rate (FDR), the false negative rate (FNR), accuracy (Acc), the F1 score, the Matthews correlation coefficient (MCC), the receiver operating characteristic curve (ROC), and the area under curve (AUC) for the evaluation model [44–46].

	Prediction Positive	Prediction Negative
Actual positive	True positive (TP)	False negative (FN)
Actual negative	False positive (FP)	True negative (TN)

True positive (TP): correctly predicting a label (predicted “yes”, and it is “yes”); True negative (TN): correctly predicting the other label (predicted “no”, and it is “no”); False positive (FP): falsely predicting a label (predicted “yes”, but it is “no”); False negative (FN): missing an incoming label (predicted “no”, but it is “yes”).

True positive rate (TPR): additionally known as sensitivity (Sn) recall, this is the proportion of samples that are judged to be positive of those that are actually positive and is calculated using the following equation:

$$TP / (TP + FN) \quad (1)$$

True negative rate (TNR): additionally known as specificity (Sp), this is the percentage of samples judged to be negative of those that are actually negative and is calculated using the following equation:

$$TN / (TN + FP) \quad (2)$$

Positive predictive value (PPV): additionally known as precision, this is a measure of what percentage of the sample that is predicted to be positive is actually positive, and the equation is as follows:

$$TP / (TP + FP) \quad (3)$$

Negative predictive value (NPV): the definition is the same meaning as for PPV, except negative (negative predictive value) instead; the equation is as follows:

$$TN / (TN + FN) \quad (4)$$

False positive rate (FPR): the number of actually negative samples that were predicted as positive as a ratio of all the actually negative samples; the equation is as follows:

$$FP / (FP + TN) \quad (5)$$

False discovery rate (FDR): the proportion of samples predicted to be positive that are actually negative; the equation is as follows:

$$FP / (TP + FN) \quad (6)$$

False negative rate (FNR): additionally known as the miss rate, this is the number of samples predicted to be negative that are actually positive as a ratio of true positive samples; the equation is as follows:

$$FN / (FN + TN) \quad (7)$$

Accuracy (Acc): the ratio of the sample correctly classified by the classifier to the total number of samples for a given test dataset, i.e., the probability of correct prediction; the equation is as follows:

$$(TP + TN) / (TP + TN + FP + FN) \quad (8)$$

F_1 score: the equation to measure the balance between true positive rate (TPR) and positive predictive value (PPV) is as follows:

$$2TP / (2TP + FP + FN) \quad (9)$$

Matthews correlation coefficient (MCC): MCC is essentially a correlation coefficient describing the correlation between the actual and predicted classifications and has values in the range $[-1, 1]$. A value of 1 indicates perfect prediction of the subject, a value of

0 indicates that the prediction is not as good as random prediction, and -1 means that the predicted classification and the actual classification do not agree at all [46]. The equation is as follows:

$$(TP \times TN - FP \times FN) / \sqrt{((TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN))} \quad (10)$$

Receiver operating characteristic curve (ROC curve): the ROC curve is a two-dimensional coordinate system where the false positive rate (FPR) is the x-axis and the true positive rate (TPR) is the y-axis.

Area under curve (AUC): in addition to the shape of the curve, the area under curve (AUC) can also be used to determine the discriminatory power of the ROC curve. The AUC value ranges from 0 to 1; the larger the value, the better. The following are the general rules for determining the AUC value [47,48]:

AUC = 0.5 (no discrimination), ROC is just diagonal.

0.7 < AUC < 0.8 (acceptable discrimination).

0.8 < AUC < 0.9 (good discrimination).

0.9 < AUC < 1.0 (excellent discrimination).

The equation is as follows, where t is FPR and ROC (t) is TPR:

$$AUC = \int_0^1 ROC(t) dt \quad (11)$$

2.10.3. Applying a Predictive Model

To predict the disease resistance of new strains with a predictive model, three strains were selected (ANT1, AB, and BB strains). The SVM results are described as follows:

\$S-fieldname: the predicted value of target field. Here, the target is death (1).

\$SP-fieldname: probability of predicted value.

\$SP-value: probability of each possible value of the flag, alive (0) or dead (1).

2.11. Detecting the Disease Resistance of New Strains

Nile tilapia (body weight: 10–80 g) were used for the challenge experiment. The experimental groups are new F_1 strains (ANT1, AB, and BB strains). To assess the effect of disease resistance, 20 fish were injected with PBS control at the same time. The fish were maintained in 30 L tanks at 28 °C after challenge. We continued observation until there were no dead fish at least for three days.

3. Results

3.1. Transcriptome Analysis

More than 50 million raw reads were assembled into 129,105 contigs (mean size: 1086.21 bp; range: 201–38,020 bp). The transcript DNA sequences were analyzed using KOG (EuKaryotic Orthologous Groups), KEGG (Kyoto Encyclopedia of Genes and Genomes), and GO ontology databases. Table S1 shows that 59,393 transcripts are annotated in KOG. A total of 11,429 transcript IDs were assigned to 377 KEGG pathways. According to the p value, Table S2 presents the top 50 KEGG pathways. In the GO annotation, 22,690 transcript IDs were annotated to 56 categories (Table S3). According to this result, seven differentially expressed transcripts related to disease resistance were found, including four hepcidin-, two progranulin-, and one piscidin-related transcripts (Table S4).

3.2. The Gene Amplification of Hamp

The gene amplification of *hepcidin/hamp* was found in the LG11 of the Nile tilapia whole-genome database by Genome Data Viewer in NCBI website, forming 12 genes, including 8 *hamp1*, 2 *hamp4*, 1 *hamp2*, and 1 *hamp3* genes. The *hamp* genes were distributed into three locations including one gene cluster containing 10 *hamp* genes (7 *hamp1*, 1 *hamp3*, and 2 *hamp4*) spanning 95.4 kb, and one individual *hamp1* gene and one *hamp2* gene in LG11 (Figure 2). These 12 *hamp* genes in LG11 encode 4 HAMPs (Figure 3).

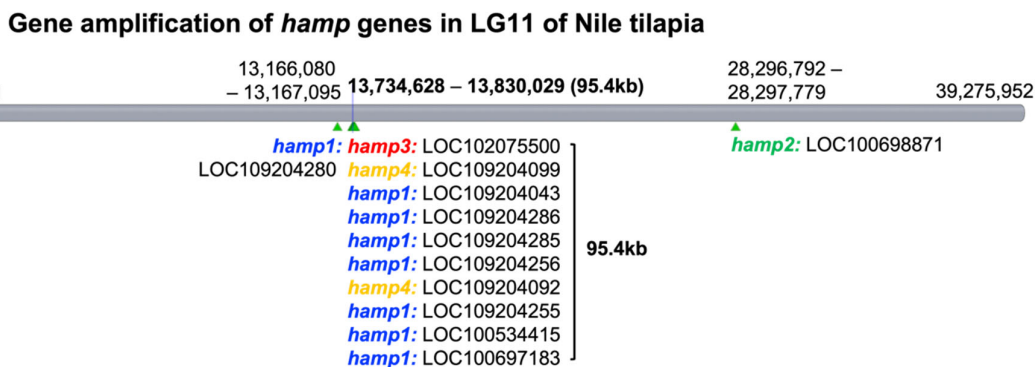


Figure 2. The gene amplification of hepcidin *hamp* genes in LG11 of Nile tilapia. There are 12 *hamp* genes including 8 *hamp1*, 1 *hamp2*, 1 *hamp3*, and 2 *hamp4* genes distributed into three locations indicated by green arrow heads in the LG11 of Nile tilapia.

(a)

```

MKAFSIAVAVTLVLAFLICMLESSAIPFAGVQEPPEEAGSNDTPVVAHQEMSAESSMMSNHIROKRQSHLSLRCWCCNCCRSNKGGCFCCRF
MKTFLTAAVTVAVVTFMCI LRSSAVPVVE-EQMLVELMNTDNPAAEPEEISVNSWKMLHNGRQKRGI FCGRCC---NGKVCWRCCYA
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSVAVAVVLTFFICVQQSSAVPVTEQEQELEEPMSMDYPAAAHEEASVDSWKMLYNSRHKRGIKCRFCCGCCTPGICGVCCRF
MKTFSAAVAVVLTFFICVQQSSAVPVTE----LEEPMSMDYPAAAEPEQLLVD SWKKLYNTIHRPGIKCRFCCGCCTPGICGICCSK
----VQLVAVAVALICICIQSSATFAEIRELQELQEAVNNDSPAAEHQEVSDSWLMPYHRQKRGIKCRFCCGCCNSGVCGLCCRF
----VQLVAVAVALICICIQSSATFAEIRELQELQEAVNNDGPAAEHQEVSDSWLTPYHRQKRGIKCCFCGCCNSGVCGLCCRF
    
```

Mature HAMP peptide

(b)

	Nile tilapia	
	HAMP4 :	GIKCRFCCGCCNSGV-CGLCCRF (22 a. a.)
Mozambique tilapia	(TH1-5) : HAMP1 :	GIKCRFCCGCCTPGI-CGVCCRF (22 a. a.)
	(TH2-3) : HAMP2 :	QSHLSLRCWCCNCCRSNKGGCFCCRF (26 a. a.)
	HAMP3 :	GIFCGRCCN----GKVCWRCCYA (19 a. a.)

Figure 3. Twelve Nile tilapia *hamp* genes in LG11 encode four HAMPs: (a) the mature HAMP peptide sequences in Nile tilapia. (b) Comparison of the HAMP mature peptides of Mozambique tilapia and Nile tilapia. The HAMP mature peptides Th1-5 and Th2-3 of Mozambique tilapia are identical to HAMP1 and HAMP2 of Nile tilapia, respectively. Moreover, HAMP3 and HAMP4 are only found in Nile tilapia.

3.3. The Expression of *Hamp* Genes Associated with Disease Resistance

NT1 Nile tilapia (body weight: 100~400 g) were selected. *S. iniae* was intraperitoneally injected into the tilapia with a lethal dose of 50 (1.4 × 10⁵ CFU/g). Liver, spleen, head kidney, gill, and brain tissues were collected at 0, 3, 6, 9, 12, and 18 h after infection for qPCR detection. In the liver, *hamp1* levels increased from 6 h and reached a maximum at 12 h, which was 46,075 times the level at 0 h (uninfected); *hamp2* levels reached a maximum at 3 h, 14.9 times higher than those at 0 h; *hamp3* levels increased from 3 h and reached a maximum at 12 h, 2741 times higher than at 0 h; *hamp4* levels increased from 6 h and reached a maximum at 12 h, 46,312 times the height of 0 h. The genes in descending order of expression were *hamp1*, *hamp3*, *hamp4*, and *hamp2* (Figure 4a). In the spleen, *hamp1* and *hamp3* levels increased from 6 h and reached a maximum at 12 h; *hamp2* levels increased from 6 h and reached a maximum at 12 h, 37 times higher than at 0 h; *hamp4* levels reached a maximum at 12 h, 290 times higher than levels at 0 h. The genes in descending order of expression were *hamp1*, *hamp2*, *hamp4*, and *hamp3* (Figure 4b). In the head kidney, *hamp1*, *hamp3*, and *hamp4* levels reached a maximum at 12 h; *hamp2* levels increased from 3 h and reached a maximum at 6 h, 14 times higher than levels at 0 h. The genes in descending

order of expression were *hamp1*, *hamp4*, *hamp3*, and *hamp2* (Figure 4c). In the gill, all the *hamp* genes' expression levels reached a maximum at 12 h. The genes in descending order of expression were *hamp1*, *hamp3*, *hamp4*, and *hamp2* (Figure 4d). In the brain, *hamp1*, *hamp3*, and *hamp4* levels increased from 6 h and reached a maximum at 18 h; *hamp2* levels increased from 3 h and reached a maximum at 12 h, 6 times higher than the levels at 0 h. The genes in descending order of expression were *hamp4*, *hamp3*, *hamp2*, and *hamp1* (Figure 4e).

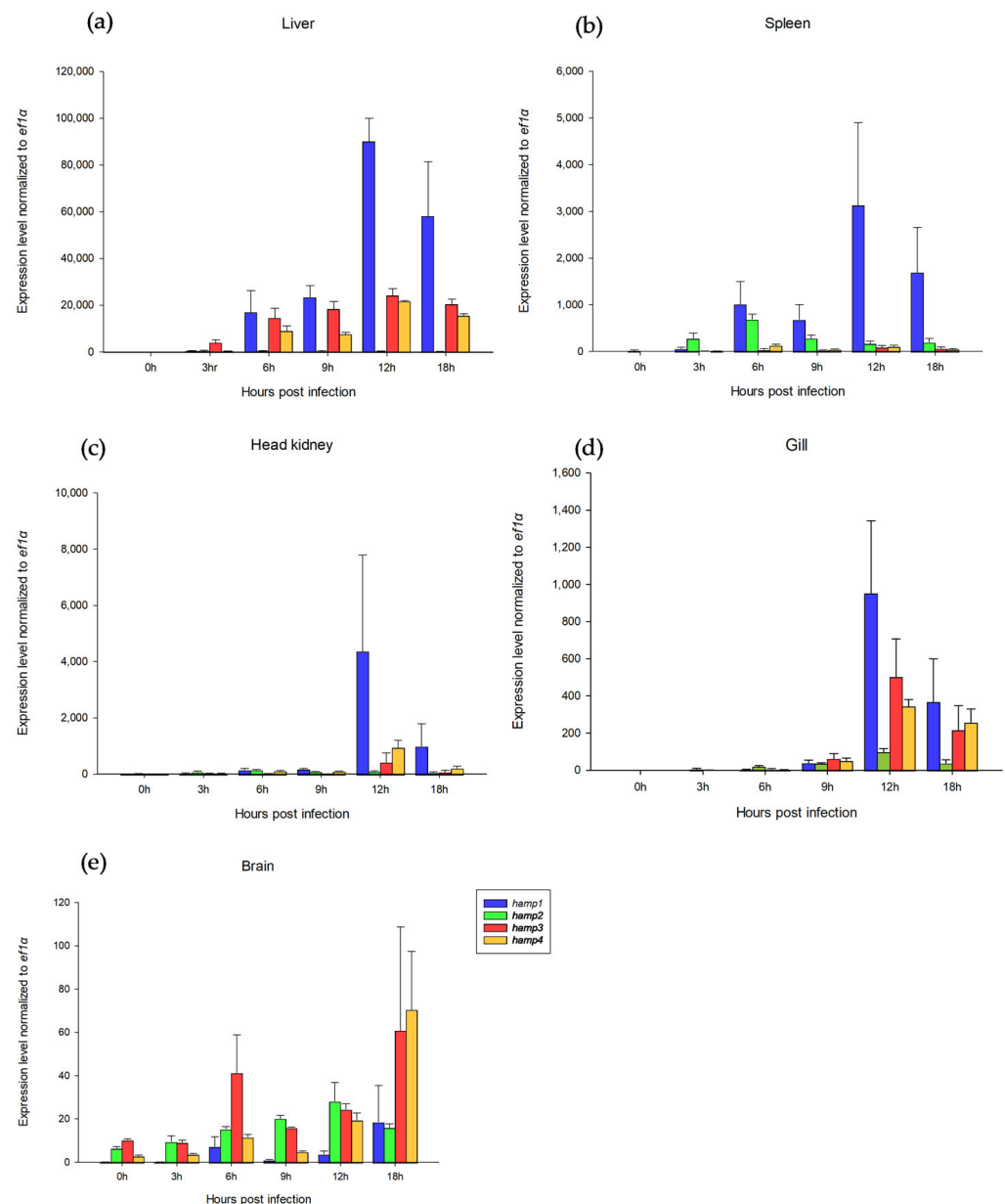


Figure 4. The gene expression of *hamp1*, *hamp2*, *hamp3*, and *hamp4* in the NT1 Nile tilapia during *Streptococcus iniae* infection (1.4×10^5 CFU/g) by qPCR. Liver, spleen, head kidney, gill, and brain tissues were collected at 0, 3, 6, 9, 12, and 18 h after infection. The gene expression was calculated by $2^{-\Delta\Delta C_T}$, and *ef1a* was the reference gene. Each bar is the mean + SEM of three independent fish: (a) the gene expression of *hamp1*, *hamp2*, *hamp3*, and *hamp4* in the liver of tilapia during *S. iniae* infection; (b) spleen; (c) head kidney; (d) gill; (e) brain.

3.4. Disease-Resistance-Associated Microsatellites

In total, we found 17 microsatellites linked to the *hamp* gene family in LG11, 4 microsatellites linked to the short type *pgrn* family in LG22, and 1 microsatellite linked to the

tilapia piscidin family in LG15 (Table 1). Table S5 shows details of the specific primers that were used.

Table 1. Disease-resistance-associated microsatellites. Total of 22 SSRs, including 17 linked to the *hamp* gene family, 4 linked to the short type *pgrn* family, and 1 linked to the tilapia *piscidin* family.

SSR	Associated Gene	SSR Name	SSR
SSR1	HAMP2 in LG11 (ID: 100698871) 5'end 3.8 kb	OnHAMP2-SSR1	(TA) ₁₁ ^a
SSR2	HAMP2 in LG11 (ID: 100698871) 5'end 3.8 kb	OnHAMP2-SSR2	(CAGG) ₆
SSR3	HAMP2 in LG11 (ID: 100698871) 5'end 3.8 kb	OnHAMP2-SSR3	(GT) ₁₃
SSR4	HAMP2 in LG11 (ID: 100698871) 3'end 4.3 kb	OnHAMP2-SSR4	(CTAC) ₇
SSR5	HAMP2 in LG11 (ID: 100698871) 3'end 4.7 kb	OnHAMP2-SSR5	(TG) ₂₁
SSR6	HAMP1 in LG11 (ID: 109204280) 5'end 2 kb	OnHAMP1a-SSR1	(GT) ₄₁
SSR7	HAMP1 in LG11 (ID: 109204280) 3'end 4.7 kb	OnHAMP1a-SSR2	(AC) ₁₂
SSR8	HAMP1 in LG11 (ID: 109204256) 5'end 1.3 kb	OnHAMP1b-SSR1	(TG) ₂₂
SSR9	HAMP1 in LG11 (ID: 109204256) 5'end 1.3 kb	OnHAMP1b-SSR2	(T) ₂₀
SSR10	HAMP1 in LG11 (ID: 109204256) 3'end 7.6 kb	OnHAMP1b-SSR3	(TG) ₁₀
SSR11	HAMP4 in LG11 (ID: 109204092) 3'end 1 kb	OnHAMP4a-SSR	(CA) ₂₉
SSR12	HAMP1 in LG11 reverse strand (ID: 109204255) 5'end 2 kb	OnHAMP1c-SSR1	(TG) ₂₇
SSR13	HAMP1 in LG11 reverse strand (ID: 109204255) 5'end 6.3 kb	OnHAMP1c-SSR2	(ATTC) ₇
SSR14	HAMP1 in LG11 reverse strand (ID: 100534415) 3'end 6.2 kb	OnHAMP1d-SSR1	(TG) ₂₂
SSR15	HAMP1 in LG11 reverse strand (ID: 100534415) 3'end 750 bp	OnHAMP1d-SSR2	(AC) ₁₃
SSR16	HAMP1 in LG11 reverse strand (ID: 100534415) intron2	OnHAMP1d-SSR3	(T) ₂₃
SSR17	HAMP1 in LG11 (ID: 109204285) 3'end 4.2 kb	OnHAMP1g-SSR	(TG) ₁₅
SSR18	PGRN2 in LG22 (ID:100692931) PGRN2 3'end 2 kb	OnPGRN-SSR1	(TG) ₁₀
SSR19	PGRN1 in LG22 (ID:100534477) PGRN1 5'end 8.2 kb	OnPGRN-SSR2	(TTGA) ₁₆
SSR20	PGRN1 in LG22 (ID:100693478) PGRN1d 5'end 1.5 kb	OnPGRN-SSR3	(A) ₄₂
SSR21	PGRN1 in LG22 (ID:100693478) PGRN1d 3'end 3.1 kb	OnPGRN-SSR4	(TGTT) ₁₀
SSR22	TP4 in LG15 (ID:100698360) TP4 5'end 5 kb	OnTP4-SSR	(GAAAA) ₆

a. The number indicates repeat numbers of individual SSR identified in the Nile tilapia genome.

3.5. Selection of Disease-Resistant Tilapia by *Streptococcus iniae* Challenge

Before selection of disease-resistant tilapia, 20 fish from the A, B, and N2 strains were infected by *S. iniae* to understand the LD (lethal dose) of each tilapia group. The results revealed that the LD70-80 of strain A was 2×10^6 CFU/g (colony forming unit per gram of body weight), strain B was 6×10^5 CFU/g, and strain N2 was 6.5×10^5 CFU/g in tilapia following *S. iniae* challenge. After LD test of each group, 200 (strain A), 198 (strain B), and 197 (strain N2) fish were challenged by *S. iniae* (6.5×10^5 to 2×10^6 CFU/g). As a result, 144 fish died in strain A, 145 fish died in strain B, and 157 fish died in strain N2 after 14 days (Table S6). In all the strains, the highest mortality was found on the first and second days, and the survival rate was stable on the last three days of the study (12–14).

3.6. The Genotype Analysis of the Disease-Resistance-Associated Microsatellites in Tilapia

The results based on the examination of 17 SSRs linked to *hamp* family genes, 4 SSRs linked to *progranulin* family genes, and 1 SSR linked to *tilapia piscidin* family gene as microsatellites associated with disease resistance in tilapia are shown in Table 1. After detection, seven *hepcidin*, three *progranulin*, and one *tilapia piscidin* family gene associated with microsatellites generated the specific product with fluorescence, including three OnHAMP2: SSR2 SSR4, SSR5; four OnHAMP1: SSR7, SSR8, SSR14, SSR17; one PGRN2: SSR18; two PGRN1: SSR20, SSR21; and one TP4: SSR22. The genotype frequency and alleles frequency of microsatellite markers were counted by GenAIEx 6.5 software (genetic analysis in Excel), and 106 alleles and 271 genotypes were found. The results of allele frequency are depicted in Table S7.

3.7. The Genetic Diversity Analysis of the Disease-Resistance-Associated Microsatellites in Tilapia

Ten disease-resistance-associated microsatellites in the A, B, and N2 strains of tilapia were detected using the GenAIEx 6.5 software, and 62, 70, 46 alleles were identified,

respectively. The mean N_a was 6.2 ± 0.512 , and no fixed (monomorphic) locus was found. Mean N_e was 3.015 ± 0.314 , lower than mean N_a ; mean I was 1.247 ± 0.104 ; mean H_o was 0.634 ± 0.051 ; mean H_e was 0.633 ± 0.039 ; mean H_o was slightly higher than mean H_e . Mean uH_e was 0.635 ± 0.039 ; mean F_{IS} was -0.042 ± 0.116 ; mean PIC was 0.583 ± 0.163 . Afterward, in the B population, mean N_a was 7 ± 1.135 , and no fixed (monomorphic) locus was found. Mean N_e was 3.907 ± 0.368 , lower than mean N_a . Mean I was 1.485 ± 0.126 ; mean H_o was 0.591 ± 0.06 ; mean H_e was 0.718 ± 0.033 ; mean H_o was lower than mean H_e . Mean uH_e was 0.72 ± 0.033 ; mean F_{IS} is 0.167 ± 0.084 ; mean PIC was 0.575 ± 0.122 . In strain N2 tilapia, mean N_a was 4.6 ± 0.653 , and no fixed (monomorphic) locus was found. Mean N_e is 2.786 ± 0.345 , lower than mean N_a . Mean I was 1.079 ± 0.125 ; mean H_o was 0.656 ± 0.095 ; mean H_e was 0.591 ± 0.048 ; mean H_o was higher than mean H_e . Mean uH_e was 0.592 ± 0.048 ; mean F_{IS} was -0.057 ± 0.109 ; mean PIC was 0.527 ± 0.170 . The results of the genetic diversity analysis are presented in Table S8. Then, three groups were compared in pairs. The results revealed that the F_{ST} of the A population to the B population was 0.114, and the Nm was 1.950; the F_{ST} of the A population to the N2 population was 0.102, and the Nm was 2.191; the F_{ST} of the B population to the N2 population was 0.098, and the Nm was 2.294. From highest to lowest, number of migrants per generation, the order of strains was N2, B, and A (Table S9).

3.8. The Correlation Analysis of Genotype and Survival Rate of the Disease-Resistance-Associated Microsatellites in Tilapia

Table 2 presents 10 of the 11 microsatellite markers that were found to have a statistically significant association ($p < 0.05$) after the correlation analysis between genotype and survival by chi-square test. The significant markers were SSR2, SSR4, SSR5, SSR8, SSR14, SSR17, SSR18, SSR19, SSR21, and SSR22. Furthermore, all correlations were at a highly significant level ($p < 0.001$), except for SSR5. However, only 1 of the 11 microsatellite markers (SSR5) was statistically significant ($p < 0.05$) after the correlation analysis by the chi-square test in strain B (Table 3). Table 4 presents 3 of the 11 statistically significant microsatellite markers ($p < 0.05$) after the chi-square test in the N2 strain, including SSR2, SSR14, and SSR22. Moreover, the correlations of SSR22 were at a highly significant level ($p < 0.001$). Furthermore, the associations between all genotypes of each SSR and the number in the alive or dead groups were analyzed by Z-test. There were four microsatellite markers (SSR4, 5, 7, 19) in the B strain and five microsatellite markers (SSR2, 14, 17, 19, 22) in the N2 strain that showed a statistically significant difference ($p < 0.05$) (Table S10).

Table 2. Correlation analysis of SSRs and survival rate by chi-square test in strain A. Total of 11 loci, 10 of the 11 SSRs (except SSR7) that are statistically significant ($p < 0.05$). Df means degrees of freedom.

Locus	Value	df	Asymptotic Significance	Count Less than 5 ^a	Minimum Expected Count
SSR2	40.463	9	<0.001	8 (40.0%)	0.32
SSR4	41.714	9	<0.001	11 (55.0%)	0.30
SSR5	31.947	16	0.010	21 (61.8%)	0.29
SSR7	2.2220	3	0.537	6 (75.0%)	0.30
SSR8	38.461	12	<0.001	16 (61.5.0%)	0.34
SSR14	45.794	7	<0.001	10 (62.5%)	0.29
SSR17	49.74	6	<0.001	8 (57.1%)	0.30
SSR18	51.019	10	<0.001	11 (50.0%)	0.28
SSR19	55.195	15	<0.001	23 (71.9%)	0.37
SSR21	74.447	23	<0.001	34 (70.8%)	0.28
SSR22	25.452	1	<0.001	0 (0.0%)	6.16

a. The genotypes with a count of less than five in both dead and alive groups. Red color = a statistically significant difference between genotype and dead or alive groups ($p < 0.05$).

Table 3. Correlation analysis of SSRs and survival rate by chi-square test in strain B. Total of 11 loci, only 1 of the 11 SSR (SSR5) that is statistically significant ($p < 0.05$). Df means degrees of freedom.

Locus	Value	df	Asymptotic Significance	Count Less than 5 ^a	Minimum Expected Count
SSR2	3.077	5	0.688	5 (41.7%)	0.19
SSR4	5.815	5	0.325	2 (16.7%)	2.93
SSR5	54.563	31	0.006	55 (85.9%)	0.28
SSR7	15.109	11	0.178	15 (62.5%)	0.28
SSR8	2.11	8	0.977	12 (66.7%)	0.26
SSR14	27.267	28	0.504	52 (89.7%)	0.27
SSR17	23.905	26	0.581	45 (83.3%)	0.28
SSR18	4.987	9	0.835	9 (45.0%)	0.30
SSR19	19.739	20	0.474	31 (73.8%)	0.33
SSR21	2.57	7	0.922	5 (31.3%)	0.27
SSR22	1.912	3	0.591	5 (62.5%)	0.28

a. The genotypes with a count of less than five in both dead and alive groups. Red color = a statistically significant difference between genotype and dead or alive groups ($p < 0.05$).

Table 4. Correlation analysis of SSRs and survival rate by chi-square test in strain N2. Total of 11 loci, 3 of the 11 SSRs (SSR2, SSR14, and SSR22) that are statistically significant ($p < 0.05$). Df means degrees of freedom.

Locus	Value	df	Asymptotic Significance	Count Less than 5 ^a	Minimum Expected Count
SSR2	21.079	4	<0.001	3 (30.0%)	1.27
SSR4	1.978	2	0.372	0 (0.0%)	5.42
SSR5	5.01	5	0.415	7 (58.3%)	0.42
SSR7	4.242	10	0.936	12 (54.5%)	0.21
SSR8	2.423	2	0.298	1 (16.7%)	3.40
SSR14	12.938	6	0.044	6 (42.9%)	0.63
SSR17	15.342	15	0.427	26 (81.3%)	0.17
SSR18	3.181	4	0.527	5 (50.0%)	0.21
SSR19	16.142	11	0.136	15 (62.5%)	0.24
SSR21	5.367	8	0.718	8 (44.4%)	0.21
SSR22	6.325	2	0.042	1 (16.7%)	3.81

a. The genotypes with a count of less than five in both dead and alive groups. Red color = a statistically significant difference between genotype and dead or alive groups ($p < 0.05$).

3.9. Predictive Model and the Effectiveness Analysis of Genotypes

For understanding the effectiveness and heritability of genotypes, predictive modeling was established for 384 fish from strains A and B (F_0). In the process of model establishment, 90% of samples were used for training and 10% of samples were used for testing, and death was used as the target (survival: 0; death: 1). Eight SSR markers were chosen for predictors (SSR2, SSR4, SSR7, SSR14, SSR18, SSR19, SSR21, and SSR22). To avoid imbalance and overfitting of data, the synthetic minority oversampling technique (SMOTE) was used as the processing method for data balancing. The SMOTE results are shown in attachment 1 (including raw training set (90%) and independent testing set (10%); adjective training set via SMOTE; predictive results).

Table 5 presents the confusion matrix of the training and testing sets. In the training set, true positive (TP) is 242, false positive (FP) is 65, true negative (TN) is 181, and false negative (FN) is 4. In the testing set, TP is 26, FP is 5, TN is 7, and FN is 1. Then, the evaluation values of the predictive model were calculated through the four parameters of TP, FP, TN, and FN (see Table 6). Figure S1 shows the receiver operating characteristic curve (ROC curve) with FPR (1-specificity) and TPR (sensitivity). Meanwhile, we calculated the area under curve (AUG). The AUG was 0.983 in the training set and 0.849 in the testing set.

Table 5. The confusion matrix of the training and testing sets. There were 492 samples in the training set. True positive (TP) is 242; true negative (TN) is 181; false positive (FP) is 4; false negative (FN) is 65. In the testing set, there was a total of 39 samples. TP is 26; TN is 7; FP is 1; FN is 5.

	Training		Testing	
	Positive Prediction	Negative Prediction	Positive Prediction	Negative Prediction
Actual positive	242	65	26	5
Actual negative	4	181	1	7

Table 6. The validation index of the training and validation sets. All validation indices were calculated through the four parameters of TP, FP, TN, and FN. In addition to the FPR, FDR, and FNR, most of the validation indices are higher than 0.7. Only NPV and MCC in the testing set are lower than 0.7.

Measure	Training	Testing	Derivations
Sensitivity	0.7883	0.8387	$TPR = TP / (TP + FN)$
Specificity	0.9891	0.8750	$SPC = TN / (FP + TN)$
Precision	0.9918	0.9630	$PPV = TP / (TP + FP)$
Negative predictive value (NPV)	0.7368	0.5833	$NPV = TN / (TN + FN)$
False positive rate (FPR)	0.0109	0.1250	$FPR = FP / (FP + TN)$
False discovery rate (FDR)	0.0082	0.0370	$FDR = FP / (FP + TP)$
False negative rate (FNR)	0.2117	0.1613	$FNR = FN / (FN + TP)$
Accuracy	0.8635	0.8462	$ACC = (TP + TN) / (P + N)$
F_1 Score	0.8784	0.8966	$F_1 = 2TP / (2TP + FP + FN)$
Matthews correlation coefficient (MCC)	0.7526	0.6244	$TP \times TN - FP \times FN / \sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}$

To determine the reliability of the predictive model, the mortality of the new *Streptococcus*-resistant group (F_1) was calculated using the predictive model. In total, 96, 55, and 40 fish were selected from strains ANT1, AB, and BB, respectively. The experimental results exhibited the predicted value and probability of predicted value (predicted probability). The results show that death was predicted for 95, 52, and 31 samples in groups ANT1, AB, and BB, respectively (Table S11). Figure 5a shows the predictive mortality of tilapia groups using the predictive model. The mean predictive mortality (\pm SEM) of ANT1 is 0.932 ± 0.0107 , AB is 0.861 ± 0.0245 , and BB is 0.765 ± 0.0414 . The differences between the ANT1 and the AB, and BB groups are significant. The new hybrid strains (F_1) were infected by *S. iniae* to further establish the prediction accuracy of the predictive models. Figure 5b presents the mortality and the IP injection dose of *S. iniae* ($2-10 \times 10^5$ CFU/g) with regression lines.

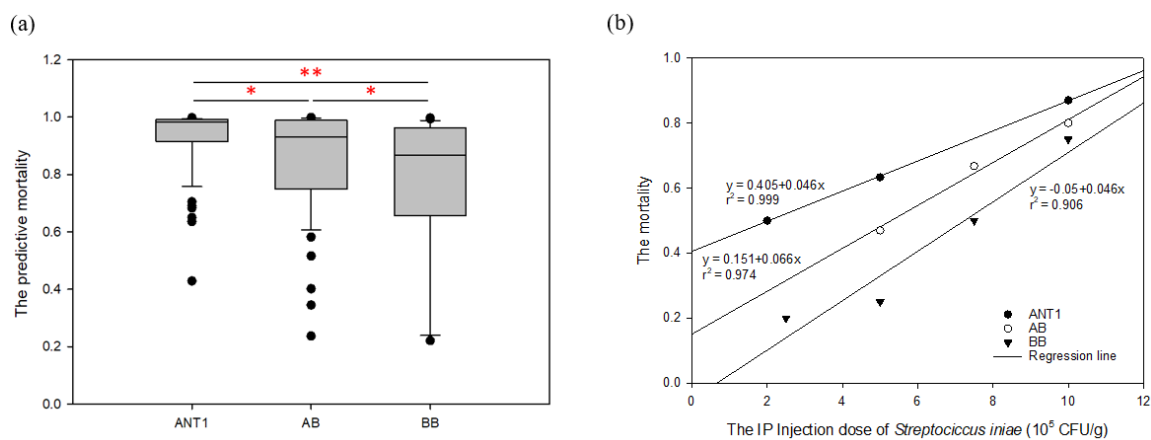


Figure 5. The predictive and actual mortality of new *Streptococcus*-resistant groups (F_1): (a) the predictive mortality of tilapia groups via the SVM predictive model. The x-axis denotes different groups, i.e., 96 fish in ANT1; 55 fish in AB; and 40 fish in the BB group. The y-axis is the predictive

mortality via the SVM predictive model; 1.0 denotes 100% mortality. The gray box represents the interquartile range (IQR); from top to bottom are upper quartile (Q3), median (Q2), and lower quartile (Q1), respectively. The upper line is maximum and the lower line is minimum; dark points denote an outlier. *: there is a statistically significant difference by one-way ANOVA and Tukey post hoc test ($p < 0.05$); **: $p < 0.01$. (b) The mortality of tilapia groups with regression lines through *S. iniae* IP injection. The x-axis is the IP injection dose of *S. iniae* (10^5 CFU/g). The y-axis is the mortality after *S. iniae* IP injection; 1.0 signifies 100% mortality.

4. Discussion

At the beginning of this study, the NT1 strain was infected by *S. iniae* 89353 (10^4 cfu/g) at 12 hpi for the collection of the transcriptome sample. In contrast to previous research [49,50], a lower dose and longer response time were selected to allow observation of the recovery response after challenging with *S. iniae*. Even if there was no enormous differential gene expression in the transcriptome result (Table S4), the qPCR results indicate that *hamp* gene expression increased substantially. Figure 4 shows the *hamp1*, *hamp3*, and *hamp4* levels reaching a maximum at 12 h (*hamp2* at 3–6 h) in all tissues during *S. iniae* infection. It was also reported that hepcidin expression increased remarkably after pathogenic infection [51,52]. Figures 2 and 3 present the results for 12 hepcidin genes in Nile tilapia. These results are similar to previous reports showing that multiple hepcidin gene copies have been generated through duplication and diversification processes in fish [19,53–55], and that Nile tilapia has more hepcidin genes than blue tilapia [18]. These results suggest that hepcidin gene amplification is associated with disease resistance in tilapia. These genes were found to be significantly upregulated after the challenge experiment. Surprisingly, hepcidin gene amplification was also observed (Figures 2 and 3). It is therefore assumed that these genes play an important role in the infective response. Furthermore, the short product of *pgrn* could enhance disease resistance via participation in the regulation of innate immune-related genes in tilapia [56]. The GRN-41 peptide, which is product of *Pgrn1* generated by alternative RNA splicing, also has antimicrobial activity against *Vibrio* [57]. The disease resistance of tilapia can be effectively increased by TP4 (tilapia piscidin 4) [58–61].

The correlation analysis by chi-square test (Tables 2–4) and Z-test (Table S10) suggest the relationship between SSRs polymorphism and disease resistance. Moreover, Figure 4 indicates the association of gene expression and disease resistance in HAMP. We hypothesized that different SSRs lengths affect gene expression, which in turn caused effects on disease resistance. The effect of SSRs on gene expression has been reported [26,27]. Some of the examples of SSRs affecting gene expression are as follows. In human, there was a long CCG trinucleotide repeat in the 5'-UTR of the *FMR1* gene. This SSR was adjacent to the promoter and affected the performance of the *FMR1* gene, leading to fragile X syndrome (FXS) [62]. In another study, the (GA)_n microsatellite sequence of the promoter region has been shown to bind the GAGA factors (GAF) proteins in *Drosophila* GAGA factor (GAF) [63,64]. The GAF is a multifunctional protein that influences gene expression, the communication between promoters and enhancers, nucleosome organization, and chromosome structure [65]. In mammalian cells, the transcription start site (TSS) at the 5'-UTR end of the promoter is affected by GAF binding sequences [66]. Streelman and Kocher present the (CA)_n microsatellite which is found in the prolactin 1 (*prl 1*) gene of the 5'-UTR is associated *prl 1* gene expression in tilapia [67]. Both *prl1* and growth hormone (GH) gene polymorphism has been proven to be linked to growth in tilapia [68]. Overall, our study is based on these disease-resistance-related genes of tilapia, and we selected samples for intraperitoneal injection of the highly pathogenic *S. iniae*. Developing microsatellites associated with these antimicrobial peptides might effectively aid in the marker-assisted selection of disease-resistant strains of aquaculture species.

In the above-mentioned study, we concluded that SSR and disease resistance are indeed associated, but the results of these markers are not similar to different strains. As the result, the correlation analyses of A, B, and N2 populations were, respectively, 10, 1, and 3 of

the 11 microsatellite markers after the correlation analysis between genotype and survival by the chi-square test ($p < 0.05$). Then, there were four microsatellite markers (SSR4, 5, 7, 19) in the B strain and five (SSR2, 14, 17, 19, 22) in the N2 strain (Table S10), both of which showed a statistically significant difference ($p < 0.05$) when analyzing the association between all genotypes of each SSR and the numbers in the alive or dead groups by Z-test. These results indicate that the disease resistance of strain A is higher than in strains B and N2, and B and N2 strains are similar. The direct reason may be the variation in disease resistance of different strains. This result is also consistent with the result of the *S. iniae* challenge experiment (the LD71.3 of strain A was 2×10^6 CFU/g, the LD73.2 of strain B was 6×10^5 CFU/g, and the LD79.7 of strain N2 was 6.5×10^5 CFU/g).

Furthermore, the correlation analysis results of all SSRs in the A, B, and N2 strains were compared. A total of 271 genotypes were found, of which 55 genotypes were related to survival (Table S10). However, many survival-related genotypes were only found in specific strains (such as SSR5, 8, and 21), even having a statistically significant difference (especially strain A, Figure 6). This may be because of the following: (1) the number of genotypes is too large, but the number of fish related to each genotype is too small; or (2) the challenge experiments with a high lethal dose create a smaller number of survivor groups. Nevertheless, the SSR5, SSR8, SSR19, and SSR22 in strain A are still significantly related to survival. It was found that the number and proportion of most genotypes that have no significant difference in other groups were still associated with survival via a comparison with the number and proportion of all microsatellite markers in the death and survival groups (Table S10). Fuji et al. [69] mentioned that, despite there being no significant correlation in the first generation of statistics, the ratio of certain genotypes could be increased by repeated backcrossing; thus, a microsatellite closely linked to lymphocystis disease resistance (LD-R) was selected from 50 microsatellites. They also attempted to transfer the closely linked LD-R microsatellite marker Poli9-8TUF into a commercial strain, and they successfully developed a new disease-resistant strain [70]. These findings were corroborated in our study; although some genotypes are only significantly associated with survival in specific strains, they are nevertheless associated with survival.

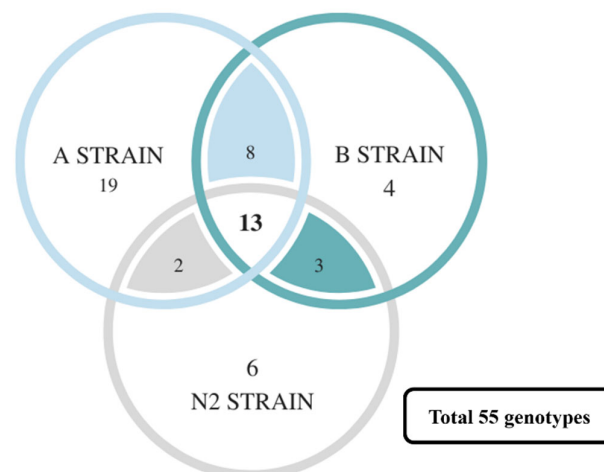


Figure 6. The Venn diagram of disease-resistance-associated genotypes in A, B, and N2 strains. Total number of genotypes is 55.

Although 55 disease-related genotypes were found by a Z-test, most genotypes were found in strain A. It was proposed that more disease-resistant markers could be found in the strain demonstrating broader disease resistance. However, those with both significant differences and no significant differences might still be potential molecular markers. As mentioned in the previous section, even if 37 genotypes only have a significant difference in strain A, the number and proportion of most genotypes which have no significant difference in other strains are still associated with survival. Moreover, disease resistance

may also be graded according to different combinations of genotypes. Our hypotheses are the following: (1) every genotype has different strengths; and (2) different combinations of genotypes cause variations in effectiveness. To address the challenges arising from this variation, predictive modeling was built by strains A and B.

Table 6 shows the evaluation values of our predictive model. Accuracy is the factor most commonly evaluated for a predictive model and is represented by a value derived from the number of samples of correct judgment (true positive and true negative) divided by all samples [71]. Accuracy was 0.8598 in the training set and 0.8462 in the testing set (range from 0 to 1; the closer to 1, the better). However, this accuracy does not apply when the actual number of positive samples is low; thus, precision and sensitivity are used. Precision and sensitivity are both concerned with a true positive, but from different perspectives. Precision relies on predicting the actual precision in a positive situation, while sensitivity predicts “how much” of the actual positive answer can be recalled in a positive situation [45]. The precision in the training set was 0.7883 and the sensitivity was 0.9837. The precision in the testing set was 0.8387 and the sensitivity was 0.9630. Both values were close to 1 (especially the sensitivity). The F_1 -score is the harmonic average of the two values (precision and sensitivity) and is often used to evaluate the accuracy of a given model [45]. The F_1 -scores were 0.8752 and 0.8966 for the training and test sets, respectively (range from 0 to 1; the closer to 1, the better). The sensitivity denotes detection of how many samples will actually die. The specificity represents the number of samples that survived via predictive model detection. The higher the sensitivity and specificity values, the better the model in terms of prediction (range from 0 to 1; the closer to 1, the better) [72]. The specificity for the training set and the test set was 0.7358 and 0.5833, respectively, which indicated that the actual surviving samples in the test set were slightly lower than in the training set. MCC is usually regarded as a balanced indicator. In essence, MMC is a correlation coefficient that describes the actual classification and the predicted classification. The range is from -1 to 1. A value of 1 describes a perfect prediction, a value of 0 shows that the predicted result is worse than the random result, and -1 demonstrates that the predicted classification and the actual classification are completely inconsistent [46]. The MCC in the training set and testing set was 0.7427 and 0.6244, respectively. Additionally, the closer the FPR, FDR, and FNR are to 0, the better. The results showed that only the FPR of the testing set was higher. This is because $FPR = 1 - \text{specificity}$, and a lower correct number of tested actual surviving samples means a higher incorrect number of tested surviving samples. Therefore, we also plotted the ROC curve (Figure S1) using false positive rate (FPR) and true positive rate (TPR). If the ROC curve is equal to the diagonal line, the model shows no discrimination. If the ROC curve moves to the upper left corner, the model is more sensitive to disease resistance (the lower false positive rate), which means the model has better discrimination. Meanwhile, the ROC curve is used to calculate the area under curve, which ranges from 0 to 1; the larger value, the better [44,45,47]. The AUC was 0.983 in the training set, which meant excellent discrimination. The AUC was 0.849 in the testing set, which meant good discrimination. Combining the evaluation values from Table 6 and Figure S1, it is proposed that this model has certain credibility in predicting the mortality of a population.

The model was applied to detect the new *Streptococcus*-resistant group (F_1). The result reveals that the proportions of predictive death numbers in the ANT1, AB, and BB groups were 95/96 (0.99), 52/55 (0.945), and 31/40 (0.775), respectively. Meanwhile, the mean predictive mortality (\pm SEM) of ANT1 was 0.932 ± 0.011 , AB was 0.861 ± 0.026 , and BB was 0.765 ± 0.041 (Table S10). Moreover, there are more samples with lower mortality in the BB group (Figure 5a). As a result, the trend of predictive disease resistance (Figure 5a) corresponded with actual disease resistance (Figure 5b). The mortality of ANT1 was highest, next was AB; BB was the lowest group. The predictive model, which was built by disease-resistance-associated microsatellites, could not only predict the mortality of a pure line but also of hybrid offspring. These results may indicate that the resistance-related genotypes which are found from F_0 are still applicable in the offspring. Moreover,

this predictive model can also estimate the mortality rate in genotype combinations. There are still numerous issues with this model: (1) the *S. iniae* dose does not ensure mortality and only an approximate relative value; (2) offspring not sharing the same genotype as their parent leads to inaccurate interpretation; and (3) the sample size may be too small to build a predictive model. Incorporating more information will increase the accuracy of predictions. To establish a precise predictive model, the dose of the challenge experiment and the time of death should be added to this SVM model, or different models should be established in further experiments. In addition, the predictive model could also be strengthened via machine learning during the breeding process.

With the development of molecular biotechnology, a huge variety of genome-based biotechnologies have been applied to the field of aquaculture research. However, most Nile tilapia breeding relies on traditional breeding methods to select phenotypes, such as growth rate, weight, and length. Relative to the aquaculture industry, modern genome-based strategies (e.g., marker-assisted selection and genomic selection breeding) have been widely using in agriculture and animal industries. Even though marker-assisted selection has only begun to be applied in the aquaculture industry in recent years, some cases of aquaculture studies can be found; for instance, high growth rate, cold resistance, and disease resistance in flatfish [69,70], rainbow trout [73–77], and carp [78,79]. Thus far, most of the current research of marker-assisted selection has focused on developing massive SNPs, SSRs, and deletions in Nile tilapia [80] for sex determination [81], population structure analysis [82], improvement of growth and fillet yield [83,84], and cold stress [85]. Often, few markers are used in breeding, which is not only time consuming but expensive. Overall, there are still only a few areas of research into disease-resistance-associated microsatellites in tilapia, but all commercial tilapia strains in Taiwan are hybridized. Previous studies may therefore not provide precise information directly relevant for the Taiwan tilapia industry.

5. Conclusions

In this study, 11 disease-resistance-associated microsatellites and 55 survival-related genotypes were identified and characterized and a predictive model for mortality linked to disease was developed. Three new *Streptococcus*-resistant strains were established through a double challenge experiment and marker-assisted selection. Even though some studies have reported that researchers successfully established *Streptococcus*-resistant tilapia, it is necessary to establish many families at the same time. Moreover, disease resistance in previous research could only be detected through challenge experiments or exposure to environmental stresses. This will be tough to use in the farms. Our findings allow selection of disease-resistant fish without gene expression. It may provide a cost-effective and time-saving strategy for assessing disease resistance and can accelerate the breeding process with the use of fewer fish, families, offspring, and markers and without the need for killing.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes13010099/s1>. Figure S1: The receiver operating characteristic curve (ROC curve) of the training and testing sets. Table S1: KOG orthologous group clustering analysis. Table S2: Distribution of the top 50 KEGG pathways and annotations of the NT1 strain. Table S3: GO annotation distribution of NT1. Table S4: The differential gene expression of *hepcidin*-, *progranulin*-, and *piscidin*-related transcripts after *S. iniae* infection. Table S5: Primer list. Table S6: *Streptococcus iniae* challenge after 14 days. Table S7: The allele frequencies of the disease-resistance-associated microsatellites. Table S8: The results of heterozygosity, *f*-statistics, and polymorphism by population for codominant data. Table S9: Pairwise population F_{ST} values and estimates of N_m . Table S10: The genotypes with count, % within group, and % within genotype in the dead and alive group. Table S11: The different combinations of genotypes and the predictive results of the new *Streptococcus*-resistant groups (F_1) through SVM predictive model. Attachment 1: The raw data and predictive results of the predictive model.

Author Contributions: Conceptualization: C.-C.C., C.-Y.L. and H.-Y.G.; methodology, C.-C.C., T.-H.H. and C.-W.H.; validation, C.-C.C., C.-H.H. and H.N.P.; formal analysis, C.-C.C., C.-H.H., H.N.P. and T.-T.L.; investigation, C.-C.C., T.-H.H., C.-W.H. and H.-Y.G.; resources, R.-H.C., S.-D.Y.,

C.-I.C. and H.-Y.G.; data curation, C.-C.C., C.-Y.L., T.-T.L. and H.-Y.G.; writing—original draft preparation, C.-C.C.; writing—review and editing, C.-C.C. and H.-Y.G.; visualization, C.-C.C.; supervision, H.-Y.G.; project administration, H.-Y.G. and C.-Y.L.; funding acquisition, H.-Y.G. and C.-Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by grant (MOST 104-2321-B-019-003-MY3) from the Ministry of Science and Technology and grants (108AS-14.3.1-ST-a2, 109AS-12.3.1-ST-a2) from the Council of Agriculture, Executive Yuan, ROC (Taiwan) to Hong-Yi Gong.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Institutional Animal Care and Use Committee of National Taiwan Ocean University (NTOU) and approved by the Council of Agriculture, Executive Yuan, ROC (Taiwan) (108AS-14.3.1-ST-a2, 109AS-12.3.1-ST-a2).

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets presented in this study can be found in online repositories. De novo transcriptome assembly of NT1 was submitted to the NCBI short read archive database (accession numbers: SRR14141863 and SRR14141864).

Acknowledgments: This research was supported by the Center of Excellence for the Oceans (National Taiwan Ocean University), which were financially supported by the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education, ROC (Taiwan). The authors would like to thank Wen-Yi Lin for helping in the collection and processing of DNA samples.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. FAO. *FAO Yearbook: Fish and Aquaculture Statistics 2015*; FAO: Rome, Italy, 2019.
2. Gjedrem, T.; Baranski, M. *Selective Breeding in Aquaculture: An Introduction*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010; Volume 10.
3. Eknath, A.E.; Tayamen, M.M.; Palada-de Vera, M.S.; Danting, J.C.; Reyes, R.A.; Dionisio, E.E.; Capili, J.B.; Bolivar, H.L.; Abella, T.A.; Circa, A.V. Genetic improvement of farmed tilapias: The growth performance of eight strains of *Oreochromis niloticus* tested in different farm environments. In *Genetics in Aquaculture*; Elsevier: Amsterdam, The Netherlands, 1993; pp. 171–188.
4. Ponzoni, R.W.; Hamzah, A.; Tan, S.; Kamaruzzaman, N. Genetic parameters and response to selection for live weight in the GIFT strain of Nile tilapia (*Oreochromis niloticus*). *Aquaculture* **2005**, *247*, 203–210. [[CrossRef](#)]
5. Amal, M.; Zamri-Saad, M. Streptococcosis in tilapia (*Oreochromis niloticus*): A review. *Pertanika J. Trop. Agric. Sci.* **2011**, *34*, 195–206.
6. Shlapobersky, M.; Sinyakov, M.S.; Katzenellenbogen, M.; Sarid, R.; Don, J.; Avtalion, R.R. Viral encephalitis of tilapia larvae: Primary characterization of a novel herpes-like virus. *Virology* **2010**, *399*, 239–247. [[CrossRef](#)]
7. Surachetpong, W.; Janetanakit, T.; Nonthabenjawan, N.; Tattiyapong, P.; Sirikanchana, K.; Amonsin, A. Outbreaks of tilapia lake virus infection, Thailand, 2015–2016. *Emerg. Infect. Dis.* **2017**, *23*, 1031. [[CrossRef](#)] [[PubMed](#)]
8. Mian, G.; Godoy, D.; Leal, C.; Yuhara, T.; Costa, G.; Figueiredo, H. Aspects of the natural history and virulence of *S. agalactiae* infection in Nile tilapia. *Vet. Microbiol.* **2009**, *136*, 180–183.
9. Noga, E.J.; Ullal, A.J.; Corrales, J.; Fernandes, J.M. Application of antimicrobial polypeptide host defenses to aquaculture: Exploitation of downregulation and upregulation responses. *Comp. Biochem. Physiol. Part D Genom. Proteom.* **2011**, *6*, 44–54. [[CrossRef](#)]
10. Smith, V.J.; Desbois, A.P.; Dyrinda, E.A. Conventional and unconventional antimicrobials from fish, marine invertebrates and microalgae. *Mar. Drugs* **2010**, *8*, 1213–1262. [[CrossRef](#)]
11. Chi, J.-R.; Liao, L.-S.; Wang, R.-G.; Jhu, C.-S.; Wu, J.-L.; Hu, S.-Y. Molecular cloning and functional characterization of the hepcidin gene from the convict cichlid (*Amatitlania nigrofasciata*) and its expression pattern in response to lipopolysaccharide challenge. *Fish Physiol. Biochem.* **2015**, *41*, 449–461. [[CrossRef](#)]
12. Smith, V.; Fernandes, J. Non-specific antimicrobial proteins of the innate system. *Fish Def.* **2009**, *1*, 241–275.
13. Rajanbabu, V.; Chen, J.-Y. Applications of antimicrobial peptides from fish and perspectives for the future. *Peptides* **2011**, *32*, 415–420. [[CrossRef](#)] [[PubMed](#)]
14. Rajanbabu, V.; Chen, J.-Y. Antiviral function of tilapia hepcidin 1–5 and its modulation of immune-related gene expressions against infectious pancreatic necrosis virus (IPNV) in Chinook salmon embryo (CHSE)-214 cells. *Fish Shellfish Immunol.* **2011**, *30*, 39–44. [[CrossRef](#)]
15. Masso-Silva, J.; Diamond, G. Antimicrobial peptides from fish. *Pharmaceutics* **2014**, *7*, 265–310. [[CrossRef](#)]

16. Wang, G.; Li, X.; Wang, Z. APD3: The antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.* **2016**, *44*, D1087–D1093. [[CrossRef](#)]
17. Cai, W.-Q.; Li, S.-F.; Ma, J.-Y. Diseases resistance of Nile tilapia (*Oreochromis niloticus*), blue tilapia (*Oreochromis aureus*) and their hybrid (female Nile tilapia × male blue tilapia) to *Aeromonas sobria*. *Aquaculture* **2004**, *229*, 79–87. [[CrossRef](#)]
18. Bian, C.; Li, J.; Lin, X.; Chen, X.; Yi, Y.; You, X.; Zhang, Y.; Lv, Y.; Shi, Q. Whole genome sequencing of the blue tilapia (*Oreochromis aureus*) provides a valuable genetic resource for biomedical research on tilapias. *Mar. Drugs* **2019**, *17*, 386. [[CrossRef](#)]
19. Mu, Y.; Huo, J.; Guan, Y.; Fan, D.; Xiao, X.; Wei, J.; Li, Q.; Mu, P.; Ao, J.; Chen, X. An improved genome assembly for *Larimichthys crocea* reveals hepcidin gene expansion with diversified regulation and function. *Commun. Biol.* **2018**, *1*, 195. [[CrossRef](#)] [[PubMed](#)]
20. Krause, A.; Neitz, S.; Mägert, H.-J.; Schulz, A.; Forssmann, W.-G.; Schulz-Knappe, P.; Adermann, K. LEAP-1, a novel highly disulfide-bonded human peptide, exhibits antimicrobial activity. *FEBS Lett.* **2000**, *480*, 147–150. [[CrossRef](#)]
21. Huang, P.-H.; Chen, J.-Y.; Kuo, C.-M. Three different hepcidins from tilapia, *Oreochromis mossambicus*: Analysis of their expressions and biological functions. *Mol. Immunol.* **2007**, *44*, 1922–1934. [[CrossRef](#)] [[PubMed](#)]
22. Pan, C.-Y.; Huang, T.-C.; Wang, Y.-D.; Yeh, Y.-C.; Hui, C.-F.; Chen, J.-Y. Oral administration of recombinant epinecidin-1 protected grouper (*Epinephelus coioides*) and zebrafish (*Danio rerio*) from *Vibrio vulnificus* infection and enhanced immune-related gene expressions. *Fish Shellfish Immunol.* **2012**, *32*, 947–957. [[CrossRef](#)]
23. Pan, C.-Y.; Peng, K.-C.; Lin, C.-H.; Chen, J.-Y. Transgenic expression of tilapia hepcidin 1-5 and shrimp chelonianin in zebrafish and their resistance to bacterial pathogens. *Fish Shellfish Immunol.* **2011**, *31*, 275–285. [[CrossRef](#)] [[PubMed](#)]
24. Zhou, J.-G.; Wei, J.-G.; Xu, D.; Cui, H.-C.; Yan, Y.; Ou-Yang, Z.-L.; Huang, X.-H.; Huang, Y.-H.; Qin, Q.-W. Molecular cloning and characterization of two novel hepcidins from orange-spotted grouper, *Epinephelus coioides*. *Fish Shellfish Immunol.* **2011**, *30*, 559–568. [[CrossRef](#)] [[PubMed](#)]
25. Phan-Aram, P.; Mahasri, G.; Kayansamruaj, P.; Amparyup, P.; Srisapoom, P. Immune Regulation, but Not Antibacterial Activity, Is a Crucial Function of Hepcidins in Resistance against Pathogenic Bacteria in Nile tilapia (*Oreochromis niloticus* Linn.). *Biomolecules* **2020**, *10*, 1132. [[CrossRef](#)]
26. Li, Y.-C.; Korol, A.B.; Fahima, T.; Nevo, E. Microsatellites within genes: Structure, function, and evolution. *Mol. Biol. Evol.* **2004**, *21*, 991–1007. [[CrossRef](#)]
27. Bagshaw, A.T. Functional mechanisms of microsatellite DNA in eukaryotic genomes. *Genome Biol. Evol.* **2017**, *9*, 2428–2443. [[CrossRef](#)]
28. Eze, F. Marker-assisted selection in fish: A review. *Asian J. Fish. Aquat. Res.* **2019**, *3*, 1–11. [[CrossRef](#)]
29. Chu, P.-Y.; Li, J.-X.; Hsu, T.-H.; Gong, H.-Y.; Lin, C.-Y.; Wang, C.-Y.; Huang, C.-W. Identification of genes related to cold tolerance and novel genetic markers for molecular breeding via transcriptome analysis in Taiwan tilapia (*Oreochromis* spp.). *Animals* **2021**, *11*, 3538. [[CrossRef](#)]
30. Chen, R.-H.; Chang, Y.-C.; Chang, K.-C.; Liu, F.-G. Comparison of growth performance of the hybridization and inbred lines in tilapia-development of fast-growing strain. *J. Taiwan Fish. Res.* **2008**, *16*, 41–47.
31. Gong, H.-Y.; Wu, S.-H.; Chen, C.-Y.; Huang, C.-W.; Lu, J.-K.; Chou, H.-Y. Complete genome sequence of *Streptococcus iniae* 89353, a virulent strain isolated from diseased tilapia in Taiwan. *Genome Announc.* **2017**, *5*, e01524-16. [[CrossRef](#)] [[PubMed](#)]
32. Eisen, M.B.; Spellman, P.T.; Brown, P.O.; Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **1998**, *95*, 14863–14868. [[CrossRef](#)] [[PubMed](#)]
33. Trapnell, C.; Pachter, L.; Salzberg, S.L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **2009**, *25*, 1105–1111. [[CrossRef](#)]
34. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357. [[CrossRef](#)] [[PubMed](#)]
35. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)]
36. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [[CrossRef](#)]
37. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)]
38. Martins, W.S.; Lucas, D.C.S.; de Souza Neves, K.F.; Bertioli, D.J. WebSat-A web software for microsatellite marker development. *Bioinformation* **2009**, *3*, 282. [[CrossRef](#)]
39. Hayden, M.; Nguyen, T.; Waterman, A.; McMichael, G.; Chalmers, K. Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. *Mol. Breed.* **2008**, *21*, 271–281. [[CrossRef](#)]
40. Serapion, J.; Kucuktas, H.; Feng, J.; Liu, Z. Bioinformatic mining of type I microsatellites from expressed sequence tags of channel catfish (*Ictalurus punctatus*). *Mar. Biotechnol.* **2004**, *6*, 364–377. [[CrossRef](#)] [[PubMed](#)]
41. Holsinger, K.E.; Weir, B.S. Genetics in geographically structured populations: Defining, estimating and interpreting F_{ST}. *Nat. Rev. Genet.* **2009**, *10*, 639–650. [[CrossRef](#)]
42. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
43. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
44. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]

45. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies* **2011**, *2*, 37–63.
46. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 1–13. [[CrossRef](#)] [[PubMed](#)]
47. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [[CrossRef](#)] [[PubMed](#)]
48. Zou, K.H.; O'Malley, A.J.; Mauri, L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* **2007**, *115*, 654–657. [[CrossRef](#)]
49. Zhu, J.; Li, C.; Ao, Q.; Tan, Y.; Luo, Y.; Guo, Y.; Lan, G.; Jiang, H.; Gan, X. Transcriptomic profiling revealed the signatures of acute immune response in tilapia (*Oreochromis niloticus*) following *Streptococcus iniae* challenge. *Fish Shellfish Immunol.* **2015**, *46*, 346–353. [[CrossRef](#)] [[PubMed](#)]
50. LaFrentz, B.R.; Lozano, C.A.; Shoemaker, C.A.; García, J.C.; Xu, D.-H.; Løvoll, M.; Rye, M. Controlled challenge experiment demonstrates substantial additive genetic variation in resistance of Nile tilapia (*Oreochromis niloticus*) to *Streptococcus iniae*. *Aquaculture* **2016**, *458*, 134–139. [[CrossRef](#)]
51. Porcheron, G.; Dozois, C.M. Interplay between iron homeostasis and virulence: Fur and RyhB as major regulators of bacterial pathogenicity. *Vet. Microbiol.* **2015**, *179*, 2–14. [[CrossRef](#)] [[PubMed](#)]
52. Jiang, X.-F.; Liu, Z.-F.; Lin, A.-F.; Xiang, L.-X.; Shao, J.-Z. Coordination of bactericidal and iron regulatory functions of hepcidin in innate antimicrobial immunity in a zebrafish model. *Sci. Rep.* **2017**, *7*, 4265. [[CrossRef](#)]
53. Yang, M.; Wang, K.-J.; Chen, J.-H.; Qu, H.-D.; Li, S.-J. Genomic organization and tissue-specific expression analysis of hepcidin-like genes from black porgy (*Acanthopagrus schlegelii* B.). *Fish Shellfish Immunol.* **2007**, *23*, 1060–1071. [[CrossRef](#)]
54. Neves, J.V.; Caldas, C.; Vieira, I.; Ramos, M.F.; Rodrigues, P.N. Multiple hepcidins in a teleost fish, *Dicentrarchus labrax*: Different hepcidins for different roles. *J. Immunol.* **2015**, *195*, 2696–2709. [[CrossRef](#)]
55. Cho, Y.S.; Lee, S.Y.; Kim, K.H.; Kim, S.K.; Kim, D.S.; Nam, Y.K. Gene structure and differential modulation of multiple rockbream (*Oplegnathus fasciatus*) hepcidin isoforms resulting from different biological stimulations. *Dev. Comp. Immunol.* **2009**, *33*, 46–58. [[CrossRef](#)]
56. Wu, S.-H.; Lin, H.-J.; Lin, W.-F.; Wu, J.-L.; Gong, H.-Y. A potent tilapia secreted granulin peptide enhances the survival of transgenic zebrafish infected by *Vibrio vulnificus* via modulation of innate immunity. *Fish Shellfish Immunol.* **2018**, *75*, 74–90. [[CrossRef](#)]
57. Wu, S.-H.; Chou, H.-Y.; Liu, P.-C.; Wu, J.-L.; Gong, H.-Y. Granulin peptide GRN-41 of Mozambique tilapia is a novel antimicrobial peptide against *Vibrio* species. *Biochem. Biophys. Res. Commun.* **2019**, *515*, 706–711. [[CrossRef](#)] [[PubMed](#)]
58. Peng, K.-C.; Lee, S.-H.; Hour, A.-L.; Pan, C.-Y.; Lee, L.-H.; Chen, J.-Y. Five different piscidins from Nile tilapia, *Oreochromis niloticus*: Analysis of their expressions and biological functions. *PLoS ONE* **2012**, *7*, e50263. [[CrossRef](#)] [[PubMed](#)]
59. Huang, H.-N.; Chan, Y.-L.; Wu, C.-J.; Chen, J.-Y. Tilapia Piscidin 4 (TP4) stimulates cell proliferation and wound closure in MRSA-Infected wounds in mice. *Mar. Drugs* **2015**, *13*, 2813–2833. [[CrossRef](#)] [[PubMed](#)]
60. Lin, W.-C.; Chang, H.-Y.; Chen, J.-Y. Electrotransfer of the tilapia piscidin 3 and tilapia piscidin 4 genes into skeletal muscle enhances the antibacterial and immunomodulatory functions of *Oreochromis niloticus*. *Fish Shellfish Immunol.* **2016**, *50*, 200–209. [[CrossRef](#)]
61. Zahran, E.; Risha, E.; Elbahnaswy, S.; Mahgoub, H.A.; El-Moaty, A.A. Tilapia piscidin 4 (TP4) enhances immune response, antioxidant activity, intestinal health and protection against *Streptococcus iniae* infection in Nile tilapia. *Aquaculture* **2019**, *513*, 734451. [[CrossRef](#)]
62. Colak, D.; Zaninovic, N.; Cohen, M.S.; Rosenwaks, Z.; Yang, W.-Y.; Gerhardt, J.; Disney, M.D.; Jaffrey, S.R. Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science* **2014**, *343*, 1002–1005. [[CrossRef](#)]
63. Tsukiyama, T.; Becker, P.B.; Wu, C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* **1994**, *367*, 525–532. [[CrossRef](#)]
64. Berger, N.; Dubreucq, B. Evolution goes GAGA: GAGA binding proteins across kingdoms. *Biochim. Biophys. Acta BBA Gene Regul. Mech.* **2012**, *1819*, 863–868. [[CrossRef](#)] [[PubMed](#)]
65. Chetverina, D.; Erokhin, M.; Schedl, P. GAGA factor: A multifunctional pioneering chromatin protein. *Cell. Mol. Life Sci.* **2021**, *78*, 4125–4141. [[CrossRef](#)] [[PubMed](#)]
66. Lee, N.; Iyer, S.S.; Mu, J.; Weissman, J.D.; Ohali, A.; Howcroft, T.K.; Lewis, B.A.; Singer, D.S. Three novel downstream promoter elements regulate MHC class I promoter activity in mammalian cells. *PLoS ONE* **2010**, *5*, e15278. [[CrossRef](#)] [[PubMed](#)]
67. Streelman, J.T.; Kocher, T.D. Microsatellite variation associated with prolactin expression and growth of salt-challenged tilapia. *Physiol. Genom.* **2002**, *9*, 1–4. [[CrossRef](#)]
68. Dias, M.A.D.; Neto, R.V.R.; de Sousa Bueno Filho, J.S.; Jaser, S.K.K.; de Assis Lago, A.; Hilsdorf, A.W.S. Growth hormone gene polymorphism associated with grow-out performance of *Oreochromis niloticus* strains. *Aquaculture* **2019**, *503*, 105–110. [[CrossRef](#)]
69. Fuji, K.; Kobayashi, K.; Hasegawa, O.; Coimbra, M.R.M.; Sakamoto, T.; Okamoto, N. Identification of a single major genetic locus controlling the resistance to lymphocystis disease in Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* **2006**, *254*, 203–210. [[CrossRef](#)]
70. Fuji, K.; Hasegawa, O.; Honda, K.; Kumasaka, K.; Sakamoto, T.; Okamoto, N. Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* **2007**, *272*, 291–295. [[CrossRef](#)]

71. Swets, J.A. Measuring the accuracy of diagnostic systems. *Science* **1988**, *240*, 1285–1293. [[CrossRef](#)]
72. Altman, D.G.; Bland, J.M. Diagnostic tests. 1: Sensitivity and specificity. *BMJ Br. Med. J.* **1994**, *308*, 1552.
73. Jackson, T.R.; Ferguson, M.M.; Danzmann, R.G.; Fishback, A.G.; Ihssen, P.E.; O’Connell, M.; Crease, T.J. Identification of two QTL influencing upper temperature tolerance in three rainbow trout (*Oncorhynchus mykiss*) half-sib families. *Heredity* **1998**, *80*, 143–151. [[CrossRef](#)]
74. Ozaki, A.; Sakamoto, T.; Khoo, S.; Nakamura, K.; Coimbra, M.; Akutsu, T.; Okamoto, N. Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Mol. Genet. Genom.* **2001**, *265*, 23–31.
75. Salem, M.; Vallejo, R.L.; Leeds, T.D.; Palti, Y.; Liu, S.; Sabbagh, A.; Rexroad, C.E., III; Yao, J. RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS ONE* **2012**, *7*, e36264. [[CrossRef](#)] [[PubMed](#)]
76. Vallejo, R.L.; Leeds, T.D.; Gao, G.; Parsons, J.E.; Martin, K.E.; Evenhuis, J.P.; Fragomeni, B.O.; Wiens, G.D.; Palti, Y. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet. Sel. Evol.* **2017**, *49*, 17. [[CrossRef](#)] [[PubMed](#)]
77. Yoshida, G.M.; Carvalheiro, R.; Rodríguez, F.H.; Lhorente, J.P.; Yáñez, J.M. Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout. *Genomics* **2019**, *111*, 127–132. [[CrossRef](#)] [[PubMed](#)]
78. Cáceres, G.; López, M.E.; Cádiz, M.I.; Yoshida, G.M.; Jedlicki, A.; Palma-Véjares, R.; Travisany, D.; Díaz-Domínguez, D.; Maass, A.; Lhorente, J.P. Fine mapping using whole-genome sequencing confirms anti-Müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *G3 Genes Genomes Genet.* **2019**, *9*, 3213–3223. [[CrossRef](#)] [[PubMed](#)]
79. Yoshida, G.M.; Barria, A.; Correa, K.; Cáceres, G.; Jedlicki, A.; Cadiz, M.I.; Lhorente, J.P.; Yáñez, J.M. Genome-wide patterns of population structure and linkage disequilibrium in farmed Nile tilapia (*Oreochromis niloticus*). *Front. Genet.* **2019**, *10*, 745. [[CrossRef](#)]
80. Yoshida, G.M.; Lhorente, J.P.; Correa, K.; Soto, J.; Salas, D.; Yáñez, J.M. Genome-wide association study and cost-efficient genomic predictions for growth and fillet yield in Nile tilapia (*Oreochromis niloticus*). *G3 Genes Genomes Genet.* **2019**, *9*, 2597–2607. [[CrossRef](#)]
81. Li, B.J.; Zhu, Z.X.; Qin, H.; Meng, Z.N.; Lin, H.R.; Xia, J.H. Genome-wide characterization of alternative splicing events and their responses to cold stress in tilapia. *Front. Genet.* **2020**, *11*, 244. [[CrossRef](#)]
82. Suebsong, W.; Poompuang, S.; Srisapoom, P.; Koonawootrittriron, S.; Luengnaruemitchai, A.; Johansen, H.; Rye, M. Selection response for *Streptococcus agalactiae* resistance in Nile tilapia *Oreochromis niloticus*. *J. Fish Dis.* **2019**, *42*, 1553–1562. [[CrossRef](#)]
83. Sun, X.; Liang, L. A genetic linkage map of common carp (*Cyprinus carpio* L.) and mapping of a locus associated with cold tolerance. *Aquaculture* **2004**, *238*, 165–172. [[CrossRef](#)]
84. Palaiokostas, C.; Kocour, M.; Prchal, M.; Houston, R.D. Accuracy of genomic evaluations of juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by sequencing. *Front. Genet.* **2018**, *9*, 82. [[CrossRef](#)] [[PubMed](#)]
85. Joshi, R.; Árnýasi, M.; Lien, S.; Gjøen, H.M.; Alvarez, A.T.; Kent, M. Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (*O. niloticus*). *Front. Genet.* **2018**, *9*, 472. [[CrossRef](#)] [[PubMed](#)]