MAPPING QTL FOR RESISTANCE TO VNN DISEASE AND ANALYSIS OF TRANSCRIPTOME IN RESPONSE TO NNV INFECTION IN ASIAN SEABASS

LIU PENG

NATIONAL UNIVERSITY OF SINAGPORE

2016

MAPPING QTL FOR RESISTANCE TO VNN DISEASE AND ANALYSIS OF TRANSCRIPTOME IN RESPONSE TO NNV INFECTION IN ASIAN SEABASS

LIU PENG

(*B.Sc., Henan Agricultural University, China*

M.Sc., China Agricultural University, China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2016

Declaration

I hereby declare that this thesis is my original

work and it has been written by me in its entirety.

I have duly acknowledged all the sources of

information which have been used in this thesis.

This thesis has also not been submitted for

any degree in any university previously.

Liu Peng

27 June 2016

ACKNOWLEDGEMENTS

My heartiest thank first goes to my supervisors, Prof. Wong Sek Man and A/P Yue Gen Hua for their excellent guidance, invaluable instructions, insightful advice and kind support throughout my PhD candidature. I really appreciate that they provided me with the opportunity to learn plenty of knowledge of molecular virology and population genetics. I am grateful to their brilliant minds and warm hearts which help me during my hard time and guided me to discover more about my future career. My truthful thanks also go to my PhD qualify examination committee and TAC members, Prof. Gong Zhiyuan, A/P Chan Woon Khiong and A/P Yin Zhongchao for their support and kind help during my PhD journey. Grateful thanks go to the National University of Singapore for awarding me the NUS research scholarship and Temasek Life Sciences Laboratory for providing continuous support.

My sincere thanks to all my colleagues: Drs. Wang Le, Ma Keyi, Bai Bin for their useful research suggestions and experienced technical help. I also thank Wan Zi Yi, Ye Bao Qing, Grace Lin, Huang Shu Qing for helping me in some of the experiments. Special thanks go to Dr. Jin Jingjing, Natascha May Thevasagayam and Prakki Sai Rama Sridatta for their bioinformatics support.

I wish to express my deepest appreciations to my family for their encouragement throughout all these years. Special thanks to my loving wife RuiMin for her continuous support, and my son HanQing for his inspiration.

List of publications and ms

1. **Liu P**, Wang L, Wan ZY, Ye BQ, Huang S, Wong SM and Yue GH. (2016) Mapping QTL for resistance against viral nervous necrosis disease in Asian seabass. ***Marine Biotechnology*** 18: 107-116.

2. **Liu P**, Wang L, Kwang J, Yue GH and Wong SM. (2016) Transcriptome analysis of genes responding to NNV infection in Asian seabass epithelial cells. ***Fish & Shellfish Immunology*** 54: 342-52.

3. **Liu P***, Wang L*, Wong SM and Yue GH. Fine mapping QTL for resistance to VNN disease using a high-density linkage map in Asian seabass (revision submitted).

4. **Liu P**, Wang L, Wong SM and Yue GH. Molecular characterization of a novel disease resistance gene *rtp3* and its association with resistance to VNN disease in Asian seabass (in preparation).

Table of Contents

Summary

Viral nervous necrosis disease (VNN), caused by nervous necrosis virus (NNV), has caused mass mortality in Asian seabass, which is an important food fish in Southeast Asia. To understand the genetic architecture and facilitate genetic improvement for VNN disease resistance in Asian seabass, quantitative trait loci (QTL) mapping for this trait was conducted. Asian seabass fingerlings of a single back-cross family were challenged with NNV at 37 days post hatching (dph). A panel of 330 mortalities and 190 surviving fingerlings was genotyped with 149 microsatellites. A linkage map consisting of 145 markers covering 24 linkage groups (LGs) was constructed. QTL analysis was conducted using interval mapping. Thirteen and 10 QTL were identified for VNN resistance and survival time, respectively. One significant QTL, spanning 3 cM in LG20, was identified for both VNN resistance and survival time, with phenotypic variation explained (PVE) of 2.2-4.1% for resistance and 2.2-3.3% for survival time, respectively. The results suggest that VNN resistance in Asian seabass is polygenetic, and it is feasible for further fine mapping.

To fine map these QTL and identify corresponding genes, 6425 SNPs from 85 dead and 94 surviving individuals using genotyping-by-sequencing were generated. A high-density linkage map were constructed, consisting of 24 LGs containing 3000 (2852 SNPs and 148 SSRs) markers, with an average interval distance of 1.28 cM. Four QTL in three LGs with PVE of 8.3 to 11.0% were identified for resistance. Another four QTL in four LGs with PVE of 7.8 to 10.9% were detected for survival time. Further dissection of the QTL with the highest PVE identified protocadherin alpha-C 2-like (*pcdhac2*) as the candidate gene. Association study on multiple families demonstrated that a six bp indel in *pcdhac2* was significantly associated with VNN disease. Furthermore, qPCR analysis showed that its expression was significantly up-regulated in the brain, muscle and skin after

NNV infection. The results could facilitate marker assisted selection for resistance to NNV in breeding programs, and lay the basis for further functional analysis of the potential corresponding gene for resistance in Asian seabass.

Transcriptome profiling after virus infection is essential for an overall understanding of the host-virus interaction. To investigate the transcriptome profiling after NNV infection, Asian seabass epithelial cells were challenged with NNV. Libraries of eight mRNA samples (6, 12, 24, 48 hours post-inoculation) of mock and NNV-challenged cells, were constructed and sequenced using RNA-seq. The transcriptome of Asian seabass consisting of 89026 transcripts with a N50 of 2617 bp was *de novo* assembled. Further analysis identified 251 differentially expressed genes (DEGs) in response to NNV infection, including receptor-transporting protein 3 (*rtp3*), *viperin*, interferon regulatory factor 3 (*irf3*) and other genes related to innate immunity. Molecular characterization of *rtp3* by qPCR, cellular localization and association study suggest that Rtp3 could play an important role in Asian seabass-NNV interaction. The data suggest that abundant and diverse genes correspond to NNV infection. This study offers vital information for identification of disease resistance genes and lay the foundation for understanding the host-virus interaction.

List of tables

List of figures

List of symbols

Virus name

BFNNV	barfin flounder nervous necrosis virus
GCRV	grass carp reovirus
IPN	infectious pancreatic necrosis
ISAV	infectious salmon anemia virus
NNV	nervous necrosis virus
OsHV-1	ostreid herpes virus type 1
PRV	piscine reovirus
RGNNV	red-spotted grouper nervous necrosis virus
SAV	salmonid alphavirus
SGIV	Singapore grouper iridovirus
SJNNV	striped jack nervous necrosis virus
TPNNV	tiger puffer nervous necrosis virus
TSV	taura syndrome virus
VHS	viral haemorrhagic septicaemia

Chemicals and reagents

dNTP	deoxynucleoside triphosphate
EB	ethidium bromide
FBS	fetal bovine serum
NaCl	sodium chloride
PBS	phosphate-buffered saline
SDS	sodium dodecyl sulfate
TBE	tris acetate electrophoresis buffer
Tris	hydroxymethyl-aminomethane

Units and measurements

°C	degree Celsius
µg	microgram
µl	microlitre
µM	micromolar
bp	base pairs
cM	centiMorgan
cm	centimeter
g	gram
h	hour
l	litre
m	molar
min	minute
ml	mililitre
mM	milimolar
nm	nanometer
ng	nanogram
rpm	revolutions per minute
s	second
u	unit
Kb	kilobases
kDa	kilodaltons

Others

aa	amino acid
AFLP	amplified fragment length polymorphism
BC	backcross
BP	biological process
CAGE	cap analysis of gene expression
CC	cellular component
CDS	coding sequence

CIM	composite interval mapping
CP	capsid protein
CPE	cytopathic effect
CRoPS	complexity reduction of polymorphic sequences
Ct	cycle threshold
DEG(s)	differentially expressed genes
DHs	double haploids
DNA	deoxyribonucleic acid
dpc	days post challenge
dph	days post hatching
EST	expressed sequence tag
FDR	false discovery rate
FPKM	fragments per kilobase of transcripts per million fragments
GBS	genotyping-by-sequencing
GEBV	genomic estimated breeding values
GFP	green fluorescent protein
GO	gene ontology
hpi	hours post-inoculation
indel	insertion-deletion
LD	linkage disequilibrium
LG	linkage group
LOD	logarithm of odds
MAS	marker-assisted selection
MAVS	mitochondria antiviral signaling
MF	molecular function
ML	maximum likelihood
MOI	multiplicity of infection
MPSS	massively parallel signature sequencing
MSG	multiplexed shotgun genotyping

NGS	next-generation sequencing
NILs	near isogonic lines
nr	non-redundant protein sequences
nt	nucleotides
ORF(s)	opening reading frame(s)
PAMP(s)	pathogen associated molecular patterns
PCR	polymerase chain reaction
PE	paired ends
PIC	polymorphic information content
ppt	parts per thousand salinity
PRRs	pattern-recognition receptors
PVE	phenotypic variation explained
qRT-PCR	quantitative reverse transcription PCR
QTL	quantitative trait loci
QTN	quantitative trait nucleotide
RAD-seq	restriction-site-associated DNA sequencing
RAPD	random amplified polymorphic DNA
RdRp	RNA-dependent RNA-polymerase
RFLP	restriction fragment length polymorphism
RILs	recombination inbred lines
RNA	ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
RRLs	reduced-representation libraries
rRNA	Ribosome RNA
SAGE	serial analysis of gene expression
SE	single end
SIM	simple interval mapping
SNP	single nucleotide polymorphism

SSRs	simple sequence repeats
TCID50	tissue culture infective dose
TLRs	toll-like receptors
UTR(s)	untranslated regions
VAMP(s)	virus associated molecular patterns
VNN	viral nervous necrosis disease

Chapter 1. Literature review

The greatest problem of humanity has always been to secure enough food to meet nutritional needs, just as an old saying goes: 'without food, nothing else matters' (Schlote 1979). The world's population is already more than 7 billion and the momentum of rapid growth shows no sign of slowing down, with a projection of reaching 9.1 billion by 2050 (Diouf 2009). This has created tremendous pressure in food production and land/water utilization. Land-based food production could reach its limitation despite the potential of further expansion and intensification. Thus, it motivates people to seek water-based food production as a supplementary food supply (Gjedrem & Baranski 2010; Gjedrem et al. 2012). The fast-expanding aquaculture industry has already provided more than 50% of seafood consumed worldwide (FAO 2014). Moreover, the food and protein this industry supplies are generally considered as healthy and nutritious due to having high content of unsaturated fat acids, and plenty of micronutrients and amino acids (Gjedrem & Baranski 2010; Gjedrem et al. 2012). Therefore, the demand for aquaculture products has been steadily growing in the past decades. This ever-growing demand coincides with the constantly decreasing availability of suitable land and water, combining with the rapidly global climate change, calls for improving the production efficiency of aquaculture while using less land and water.

To meet this challenge, improving the performance of the relevant traits in aquaculture species is the key. This can be achieved by conventional selective breeding programs, which were first introduced in plants in 1900 and livestock in 1915 (Gjedrem & Baranski 2010). This conventional breeding approach typically involves endless cycles of crossing of elite individuals, performance evaluation of the offspring population, selection of elite offspring and crossing the elite offspring again (Gjedrem & Baranski 2010). This

process is conducted in a species over generations, resulting in gradual improvement in performance if the traits are genetically inherited. This approach of conventional selective breeding has been extensively proven to be a huge success in domesticated animals (Gjedrem & Baranski 2010). Despite the practical improvement of performance and the resulting huge economic benefit in plants and domesticated animals, the selective breeding in aquatic species is lagging far behind. In 1975, family based selective breeding was introduced in Atlantic salmon (*Salmo salar*), leading to 8.2% improvement in production until 2010 (Gjedrem & Robinson 2014). Since then, selective breeding has been initiated in several other aquatic species, mainly focusing on growth-related traits (Gjedrem & Baranski 2010).

In spite of these progress, in aquaculture, one of the major restricting factors is disease infection, thus traits of disease resistance are extremely important. Major reasons of disease being a serious issue in aquaculture include, minimal opportunity to escape or avoid each other in intensive captive culture systems, and incidental interactions between fish and pathogens under natural conditions often become fatal infections due to additional stress from biological, physical and chemical factors (Wedemeyer 1996). In addition, reduced genetic diversity in aquaculture species derived from wild strains is also a major contributor to the outbreak of disease, as such species is not given sufficient time to adapt to new disease pressure in the culture environment (Duarte et al. 2007). Using vaccines against virus infection has been long regarded as a promising way for better management of disease control. For example, a vaccine against infectious pancreatic necrosis (IPN) has been developed and implemented in Atlantic salmon; however, the protection is variable and incomplete (Mikalsen et al. 2004). Another alternative method to combat viral diseases is to use disease resistant lines, which is long regarded as the most effective, efficient and economical. However, similar to other important economic traits, disease resistance is a

quantitative trait and controlled by quantitative trait loci (QTL). Therefore, identification of QTL is an essential step towards improving the disease resistance in a selective breeding program. In recent years, selective breeding programs for disease resistance have been initiated in several aquaculture species (Gjedrem & Baranski 2010; Ødegård *et al*. 2011a; Yue 2014).

Asian seabass (*Lates calcarifer* Bloch 1790) is an important food fish in Southeast Asia and Australia, with annual production reaching 77538 tons in 2013 (FAO 2016). Despite the increasing economic importance of Asian seabass to these regions, this fish species suffered from severe nervous necrosis virus (NNV) infections during the rapid expansion of the Asian seabass industry. Facing the increasing threats from VNN infection, mapping QTL for disease resistance in Asian seabass would provide a possible solution to this problem. Genetic markers tightly linked with QTL for disease resistance could be used to select disease resistance elites and thus improves disease resistance in a population level. Besides mapping the QTL for resistance to viral nervous necrosis (VNN) disease, transcriptome analysis could strength our understanding of Asian seabass-NNV interaction. The transcriptome profiling in response to virus infection, in turn, could provide vital information for identification of disease resistance genes. Hence, mapping QTL for resistance to VNN disease and investigation of the transcriptome in response to NNV infection could have the potential to increase disease resistance in Asian seabass. The following sections form the literature review.

1.1. Asian seabass and its aquaculture production

1.1.1. Taxonomy, distribution and diversity

Asian seabass is a catadromous fish species belonging to the family *Latidae* in the order *Perciformes* (Nelson 2006). This family contains three extant genera, *Hypopteurs*,

Psammoperaca and *Lates* (Mooi & Gill 1995; Otero 2004). The genus of *Lates* consists of 15 extant species. Most of them are constrained to fresh and brackish waters in Africa's tropical areas (Otero 2004; Pethiyagoda & Gill 2012). On the other hand, these *Lates* distributed in the Indo-West Pacific region, from the Gulf of Persia to Southeast Asia with extension to Northern Australia, are considered to comprise only one species *L. calcarifer* (Mathew 2009; Jerry 2013) (Fig. 1.1). It has 75 local names and is commonly known as Asian seabass and barramundi in Australia. Just like the diversity of its names, recent studies have revealed that a considerable number of genetic variations are present in the population of Australia (Salini & Shaklee 1988; Keenan & Salini 1990). More recently, a study, using mitochondrial DNA and microsatellites, has shown that there is an even larger genetic distance between stocks in Australia and these in Southeast Asia (Chenoweth *et al.* 1998; Yue *et al.* 2009). The latest studies, using mitochondrial cytochrome c oxidase I gene and supplemented by morphological analysis, have demonstrated that Asian seabass from Australia, Myanmar (*Lates uwisara*) and Sir Lanka (*Lates lakdiva*) could be classified into three different species (Ward *et al.* 2008; Pethiyagoda & Gill 2012).

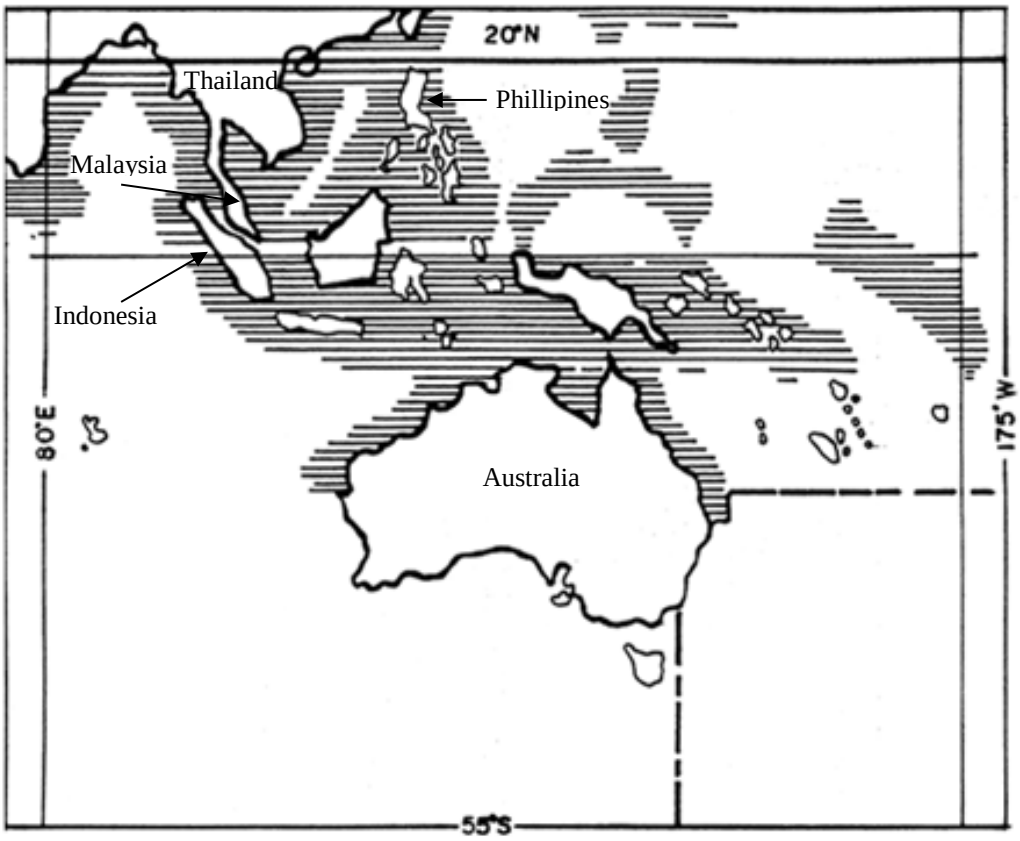


Fig. 1.1 The geographic distribution of Asian seabass. Downloaded from FAO-Training manual: Biology and culture of seabass (*Lates calcarifer*) http://www.fao.org/docrep/field/003/ac230e/ac230e02.htm

1.1.2. Commercial importance as an aquaculture food fish species

Asian seabass is an important food fish in both Southeast Asia and Australia. Since the aquaculture commencement of this species in 1970s, its production has been growing steadily and annual global production reached 77538 tons in 2013 (Fig. 1.2) (FAO 2016). The major producers are from Southeast Asia including Malaysia, Thailand and Indonesia, as well as Taiwan and Australia (FAO 2016). It is a warm water species and can only be cultured in area where water temperature remains above 25°C. Moreover, it is an euryhaline species and highly tolerant to a wide range of salinities from fresh water, to brackish and full marine water. The successful cultivation and rapid expansion of this species have been attributed to its high fecundity, development of large-scale hatching and

nursery techniques including live feed containing omega-3 fatty acids which are critical in the pre-weaning stage (Dhert *et al.* 1990; Dhert *et al.* 1992). In addition, due to its mild flavored white flesh, it has been advocated to be the 'next big fish' by industrial-scale farmers (Pierce 2006). Furthermore, it contains relatively high level of omega-3 fatty acids (Xia *et al.* 2014), which is also a reason contributing to its popularity.

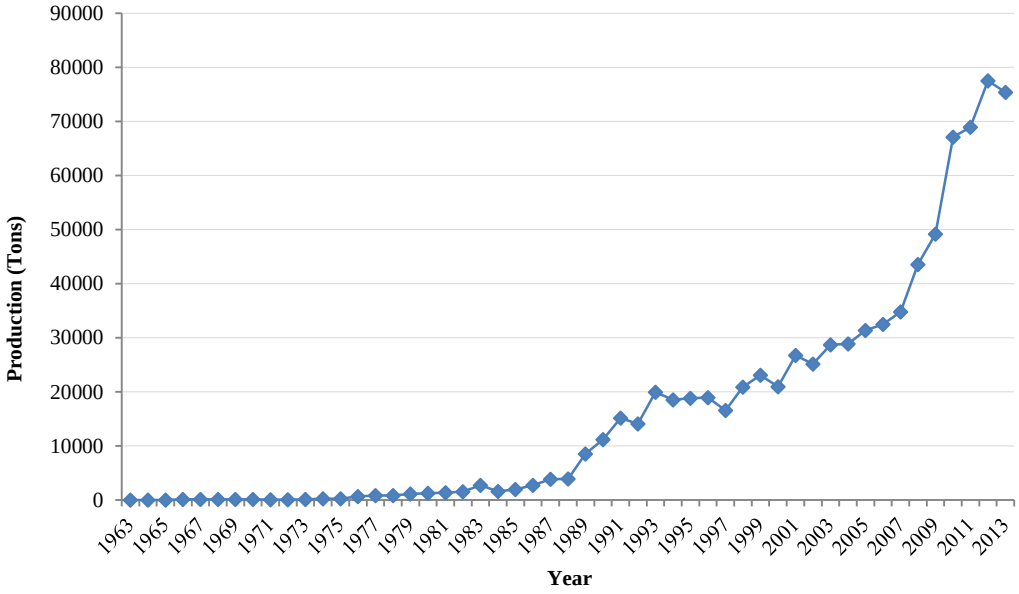


Fig. 1.2 Global production of Asian seabass. Data downloaded from FAO-Fisheries and Aquaculture Information and Statistics Service-accessed on 20/04/2016.

The successful spawning of captive Asian seabass in Thailand in 1971 has produced a far-reaching effect for its whole industry (Tookwinas 1989). The gonad maturation, spawning and larvae rearing has to be carried out in sea water (Barlow 1997). Fertilized eggs can finish hatching within a day. Fry depend on live feed such as rotifers at 2 days post hatching (dph) and artemia from 14 dph. Artificial feed can be added at 20 dph for weaning (Schipp *et al.* 2007). It is worth noting that Asian seabass is a cannibalistic species. Therefore, grading must be carried out from 26 to 60 dph every three to four days to control cannibalism and thus avoid heavy loss (Ribeiro & Qin 2013).

1.2. Infectious diseases in Asian seabass

1.2.1. Disease outbreaks result from dynamic interactions among host, pathogen and environment

Despite the continuously growing production of Asian seabass, it has been severely threatened by various infectious diseases including viral, bacterial, fungal and parasite infection (Jerry 2013). Actually, disease is the largest single factor causing huge economic loss in the aquaculture industry. Disease outbreaks are a result of dynamic interactions among the factors of host, pathogen and aquatic environment. Therefore, manipulation of these factors can either result in an outbreak or prevent disease (Jerry 2013).

1.2.2. Host factors

Fish is among the first animals to acquire both innate and adaptive immune response through evolution and thus stand at the crossroad between innate and adaptive immunity. It is well known that adaptive immunity is a strong immune response and responsible for long lasting immune memory. In addition, vaccination could take advantage of this to boost immune response before pathogen infection (Magnadottir 2010; Uribe *et al.* 2011). However, it is also known that adaptive immunity is developed much later in marine fish species than fresh water species, which prevents fish from using adaptive immunity at the early age (Magnadottir 2010; Uribe *et al.* 2011). Thus, innate immunity is critical to fish in combating with disease infection especially for the young fish. Identification of genes involved in innate immunity through transcriptome analysis or QTL mapping are important for potential improvement of disease resistance. Furthermore, host responses to infection include persistent or acute inflammatory responses. Along with the pathogen itself, these responses may cause damage to the host, as tissue degeneration and

necrosis are often observed in the infected individuals. These damages could translate into the clinical signs and symptoms. As a consequence, besides taking measures to eliminate causative agents from the host, good disease management measures should also support tissue healing (Roberts 2012).

1.2.3. Pathogen factors

Fish are consistently in contact with aquatic environments which are enriched in various microorganisms including pathogens. The interaction between host and pathogen could become fetal under certain conditions (Jerry 2013). Despite the great improvement of hatchery and culture production, Asian seabass has still been threatened by dozens of viral, bacterial, fungal and parasitic diseases. Major viral agents are NNV from family Nodaviridae, and iridovirus from family Iridoviridae, causing VNN and lymphocystis diseases, respectively. Both diseases could cause up to 100% mortality of Asian seabass. Most viruses are too small to be seen from optical microscope and are spread through vectors, via the fecal-oral route, physical contact and contained food or water (Jerry 2013). Bacterial species, including those from genera *Steptococcus*, *Vibrio*, *Flexibacter* and *Epitheliocystis*, are the major causative agents for several important bacterial diseases. Bacteria are typically a few μ m in length, with varying shapes from spheres, rods to spirals. They could typically cause up to 70% mortalities in infected Asian seabass (Jerry 2013).

Most viral and bacterial pathogens reported in the diseased fish are commonly present in the aquatic environment, thus posing consistent threats to the fish host. Hence, it would be logical to remove all the potential pathogens or reduce the pathogen loads to a certain level in the culturing environment. Moreover, it would be also beneficial to remove infected fish in time to minimize further transmission of disease. In some cases, many opportunistic pathogens could interact with the host under normal conditions without

causing diseases, but become virulent and result in severe infection when the host immunity is suppressed under stress conditions. As a consequence, avoiding causing too much stress during transporting, water changing, handing should reduce the incidence of disease outbreaks (Roberts 2012).

1.2.4. Aquatic environment

Aquaculture may be practiced in closed ponds, tanks with recirculation systems and sea cages in the open sea with intensive or semi-intensive. In these systems, fish waste or unconsumed feed could promote the building up of nitrogen and ammonia, which could produce severe negative effects on the fish. Therefore, increasing the water exchange rate to keep nitrogenous waste at a low level is necessary for fish disease management (Brown 1993). Furthermore, accurate records of the amount of feed and number of fish stocks could also be a positive contribution (Brown 1993). In addition, water quality, dissolved oxygen level, temperature and pH value are also important factors for disease outbreaks (Jerry 2013). Thus, taking effective measures to balance all those factors could greatly decrease the chance of disease outbreaks.

1.3. Viral nervous necrosis disease

1.3.1. Distribution of VNN disease

VNN disease, also known as encephalomyelitis (Bloch *et al.* 1991) and vacuolating encephalopathy and retinopathy (VER) (Munday *et al.* 1992), was first described in Asian seabass in Australia in 1987 (Glazebrook & Campbell 1987). In the following year, a similar description was also reported in European seabass (*Dicentrarchus labrax*) in the Caribbean (Bellance & de Saint-Aurin 1988). Later, it was reported in Japanese parrotfish (*Oplegnathus fasciatus*) (Yoshikoshi & Inoue 1990), turbot (*Scophthalmus maximus*)

(Bloch *et al.* 1991), red-spotted grouper (*Epinephelus akaara*) (Mori K *et al.* 1991) and striped jack (*Pseudocaranx dentex*) (Mori *et al.* 1992). Currently, it has been reported in more than 40 (22 families and 8 orders) marine and freshwater species distributed from warm to cold water areas (Shetty *et al.* 2012). It is a world-wide disease and is documented in all continents where intensive culture practice is rampant, except South America (Shetty *et al.* 2012). It is regarded as one of the most devastating infectious diseases, posing a severe threat to both mariculture and the wild population, as it has a wide host range and high infectivity (Panzarin *et al.* 2012).

1.3.2. Nervous necrosis virus

In 1990, a study reported that the causative agent of VNN disease in Japanese parrotfish was a non-enveloped and spherical virus particle of 34 nm in diameter, in the brain and eye tissues (Yoshikoshi & Inoue 1990). In the same year, a similar virus particle with a diameter of 25-30 nm was also reported in the in the brain and retina of diseased Asian seabass larvae (Glazebrook *et al.* 1990). Later, the similar virus particle was reported in more species in a number of countries (Shetty *et al.* 2012).

The causative agent of VNN disease is NNV, also known as betanodavirus, belonging to the family Nodaviridae. It contains a bipartite genome consisting of two positive-sense single-strand RNA (+ssRNA) molecules lacking poly (A) tails at their 3'-ends (Fig. 1.3). The larger segment RNA1 is about 3103 nt with a G+C content of 49.6%, and contains an open reading frame (ORF) of 2949 nt flanked by a 78 nt of 5'-untranslated region (UTR) and a 77 nt 3'-UTR. It encodes protein A of 982 amino acid (aa), a structural protein with a deduced molecular mass of 111 kDa. Protein A is a RNA-dependent RNA-polymerase (RdRp), which locates in mitochondria and is responsible for viral genome replication. The smaller segment RNA2 is about 1433 nt, with a G+C content of 53.24%,

and contains an ORF of 1017 nt flanked by a 26 nt 5'-UTR and a 390 nt 3'-UTR. It encodes another structural protein α of 338 aa, with a molecular mass of 42 kDa. The primary protein α undergoes a self splicing to produce protein β, the mature capsid protein (CP) which is used to form the virus particles with the RNA genome, and protein γ. During the replication, it also produces a sub-genomic RNA3 derived from the 3' end of RNA1. RNA3 is 371 nt with a G+C content of 62%. It contains two overlapping ORFs encoding two non-structural proteins, protein B1 (111 aa, 11 kDa) and B2 (75 aa, 8.5 kDa) (Johnson *et al.* 2001; Nakai *et al.* 2009; Shetty *et al.* 2012). Protein B1 has a function of inhibiting apoptosis at the early stage of infection (Chen *et al.* 2009). Protein B2 is a RNAi suppressor with a function of binding to the double strand RNA replicates to protect it from dicer-mediated cleavage (Fenner *et al.* 2006b; Fenner *et al.* 2006c; Fenner *et al.* 2007) (Fig. 1.3).

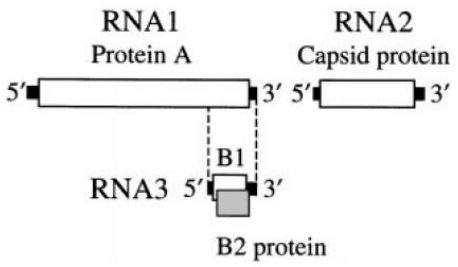


Fig. 1.3 Structural organization of the NNV genome and encoded proteins. Reprinted from Johnson and his team (Johnson *et al.* 2001).

According to their genome sequences, NNV is classified into four groups, designated as red-spotted grouper NNV (RGNNV), striped jack NNV (SJNNV), barfin flounder NNV (BFNNV) and tiger puffer NNV (TPNNV) (Shetty *et al.* 2012) (Table 1.1). The different groups have different host range and growth temperatures. Among them, RGNNV is the most common one and causes severe diseases in a broad range of warm water fish species, especially groupers and Asian seabass (Shetty *et al.* 2012).

Table 1.1 Genotype and phenotype variations of betanodavirus (NNV). Reprint from Nakai and his team (Nakai *et al.* 2009).

Genotype	Serotype	Major fish host	Optimum in vitro growth temperature
Striped jack nervous necrosis virus (SJNNV)	A	Striped jack	20-25°C
Tiger puffer nervous necrosis virus (TPNNV)	B	Tiger puffer	20°C
Redspotted grouper nervous necrosis virus (RGNNV)	C	Asian seabass, European seabass Redspotted grouper, Other groupers	25-30°C
Barfin flounder nervous necrosis virus (BFNNV)	C	Atlantic cod, Atlantic halibut Barfin flounder, Japanese flounder	15-20°C

1.3.3. VNN disease in Asian seabass

With the rapid expansion of the Asian seabass industry, frequent VNN outbreaks are reported in almost all the places with cultured Asian seabass (Jerry 2013). Asian seabass is extremely susceptible to VNN especially at the larvae stage, with more than 90%, even up 100% mortality, and continues to be threaten at the juvenile and market size adult stages, resulting in huge economic losses (Shetty *et al.* 2012; Jerry 2013). Furthermore, both vertical and horizontal transmissions of this disease have been implicated for the outbreaks. Vertical transmission of NNV could occur through reproductive fluids of both male and female brooders and fertilized eggs (Maeno *et al.* 2004; Azad *et al.* 2005; Azad *et al.* 2006; Parameswaran *et al.* 2008; Ransangan & Manin 2010). Horizontal transmission could be through cohabitation with diseased fish, feeding with contaminated fish and supplying contaminated water (Hick *et al.* 2011; Manin & Ransangan 2011). Therefore, it is considered as the most devastating disease in the Asian seabass industry. The clinical signs of infection include loss of appetite, laying down on rest, abnormal swimming patterns of 'whirling' and head down, inflation of bladders, and dark coloration in the body. Dissection of diseased fish shows that cellular vacuolation and tissue degeneration occur in the brain

and retina, which implies that these are the main organ and tissue attacked by NNV (Shetty *et al.* 2012; Jerry 2013).

1.4. Mapping quantitative trait loci (QTL) for disease resistance in aquatic species

Disease has generated numerable problems, including the major one of mortality, downgrading meat quality at slaughter and antibiotic pollution. Therefore, it is considered as one of the most restricting factors affecting aquaculture. Disease resistance, like many other traits of economic importance, is a complex/quantitative trait, which implies that it is controlled by many genes with small effects, influenced by environment and gene-environment interaction (Gjedrem & Baranski 2010; Ødegård *et al.* 2011a). QTL mapping is procedure to identify chromosome regions or genes controlling quantitative traits using DNA markers evenly covering the whole genome (Geldermann 1975). It consisted of several steps, including establishing a mapping population, phenotyping the mapping population, genetic marker discovery and genotyping the mapping population, constructing a linkage map, and linking genotype to phenotype through statistical analysis. I will present each step in the following sections.

1.4.1. Establishing a mapping population

Generating a mapping population is the first step to linkage and QTL mapping. In self-pollinating plant species, mapping populations generally originate from two highly homozygous parents with divergent phenotypes (Collard *et al.* 2005). Unlike plant species, aquatic species are commonly out-breeding species, which means that they generally do not tolerate high inbreeding. Therefore, there is usually a lack of ideal mapping families, including recombination inbred lines (RILs), near isogonic lines (NILs) and double

haploids (DHs), in aquatic species (Gjedrem & Baranski 2010). Nonetheless, the high fecundity of most aquatic species could produce large family sizes from a few parents, which could compensate for the disadvantage of lacking ideal mapping families. In addition, due to their out-breeding nature, aquatic species are generally considered as highly heterozygous in the natural population, which could directly provide good mapping resources without special cross design. That, in turn, becomes an advantage of most aquatic species (Gjedrem & Baranski 2010). As long as there are segregations of genotypes and phenotypes in the mapping population, F1 and F2 (derived from hybrid F1) populations generated from genetically diverse or highly heterozygous brooders are suitable for QTL mapping in aquaculture species (Yue 2014). Beside F1 and F2 population, backcross (BC) populations (derived from crossing F1 to one parent) are widely used for linkage mapping in aquaculture species (Yue 2014). This approach is frequently employed to introduce a specific trait, such as disease resistance, to an elite line (Xu 2010). It is worth noting that even if the trait values of parents were unknown, it would still be feasible to use their descent populations (F1, F2 and BC) to conduct QTL mapping because the reassortment and recombination of different alleles in the offspring population could produce a range of phenotypic values (Miles & Wayne 2008). Besides the type of mapping population, the size of mapping population is another important factor affecting the performance of QTL mapping. Primary QTL mapping generally requires 50-250 individuals (Mohan *et al.* 1997). Fine mapping prefers a larger size population (>500) because more recombinants could be generated (Collard *et al.* 2005; Yue 2014).

1.4.2. Marker discovery and genotyping

Genetic markers are essential for construction of linkage maps. With the rapid advances in DNA sequencing and genotyping technology, various genetic markers, each

having advantages and disadvantages have been developed and quickly applied in aquatic species (Liu & Cordes 2004; Gjedrem & Baranski 2010; Yue 2014).

1.4.2.1. Conventional genetic marker discovery and genotyping

Currently, there are several widely used conventional markers in aquatic species. Among them, allozymes, which are the protein products of a single gene locus, are the earliest markers used in aquatic species. Genotype scoring can be recorded on a starch gel in an electrical field. However, the limited number of available allozyme loci prevents them from being used in large-scale genotyping (Liu & Cordes 2004; Gjedrem & Baranski 2010). Restriction fragment length polymorphisms (RFLPs), considered as the hallmark of the genome revolution, employs restriction enzymes to digest genomic DNA, and variable lengths of products can be differentiated through gel electrophoresis. Nevertheless, the low level of polymorphisms is the main restricting factor precluding them from large-scale use (Liu & Cordes 2004; Gjedrem & Baranski 2010). Taking advantage of PCR technology, random amplified polymorphic DNA (RAPD) was developed to randomly amplify certain regions of genomic DNA with primer pairs of 8-10 bp. Its major disadvantages include low reproducibility and inability to demonstrate Mendelian inheritance (Liu & Cordes 2004; Gjedrem & Baranski 2010). To overcome the disadvantages of both RFLP and RAPD, amplified fragment length polymorphisms (AFLPs) was developed to combine their strengths while minimizing their disadvantages. This is also a method based on restriction enzyme digestion of genomic DNA and subsequent PCR to produce fragment polymorphisms, which are separated in gel (Liu & Cordes 2004; Gjedrem & Baranski 2010). This is a relatively powerful tool to produce a large number of polymorphisms with a high reproducibility. For example, 23 out of 122 AFLP markers were used to distinguish coastal cutthroat trout (*Oncorhynchus clarkii* clarkii), rainbow trout (*Oncorhynchus mykiss*) and their hybrids (Young *et al.* 2001). Nevertheless, just like the other above-

mentioned technologies, it also suffers from several major shortcomings, including special requirement for gel electrophoresis and staining, which greatly reduces the throughput for large-scale genotyping (Liu & Cordes 2004; Gjedrem & Baranski 2010). With such inherent limitations, these technologies gradually lost their attraction in many aquatic species.

Microsatellites consist of variable copies of tandemly arranged simple sequence repeats (SSRs) with a repeat number from 2-6 bp. The discovery of microsatellites could provide an opportunity to genotype a larger population with high-throughput and relatively low cost (Liu & Cordes 2004; Gjedrem & Baranski 2010; Yue 2014). Microsatellites have several features, which make them extremely suitable for genetic marker based analysis. Microsatellites are abundant across all species, with an estimated one in every 10 kb in fish species. They also appear to be evenly distributed across all chromosomes in the genome. More importantly, a substantial number of microsatellites are found in gene coding regions, including UTRs, exons and introns, as well as in the regulatory regions of genes, including promoters, enhancers and silencers. These microsatellites could be functional because they could directly or indirectly influence gene transcription, splicing, transcript satiability, degradation and transportation, translation efficiency, even changing of the protein product, and eventually the expression of traits (Li *et al.* 2004; Liu & Cordes 2004; Gjedrem & Baranski 2010; Yue 2014). In addition, microsatellites are multi-allelic due to the variable number of repeats and thus have the highest polymorphic information content (PIC) compared with any other genetic marker. Furthermore, microsatellites follow Mendelian inheritance and are co-dominant markers. However, up-front development of microsatellites requires a large amount of investment, time and effort, as the genomic segment containing microsatellites should be enriched and flanking regions of microsatellites must be sequenced in order to design targeting primer pairs. Once

identified, primer pairs can be applied throughout the population in the same species and can even be transferred to close species, thus greatly reducing the development cost. Moreover, introducing fluorescent labeled primers and automated scoring software like GeneMapper significantly increases the throughput. With the advantages of being highly polymorphic, highly repeatable, co-dominant, easily and accurately scored, microsatellites have become extremely popular in many genetic investigations in a variety of fish species, including parentage assignment, kinship mapping, population structure, product traceability and genome mapping (Liu & Cordes 2004; Gjedrem & Baranski 2010; Yue 2014).

The most abundant polymorphism in any genome is single nucleotide polymorphism (SNP), which describes the DNA sequence variation caused by point mutation in a single locus (Sachidanandam *et al.* 2001). Theoretically, all the four nucleotides of A T C and G could occur at any SNP locus, and thus could produce four alleles. Practically, most SNPs are restricted to either two pyrimidines C/T or two purines A/G. In addition, it is bi-allelic with moderate PIC, which could offset the great abundance of SNPs in the genome. Obviously, they are inherited as co-dominant markers. It is worth noting that SNPs can be found anywhere in the genome, including both non-coding and coding regions. Just like microsatellites, SNPs in the gene regulatory region, including enhancer/silencer and promoter, could affect gene transcription. SNPs in gene regions, including UTRs, introns and exons, could influence transcription, splicing, transcript stability and degradation and translation, which are closely related to trait expression. SNPs in the exon could be synonymous/silent mutations, which do not alter the sequence of aa due to degeneration of genetic code. More importantly, SNPs in the exon could also be nonsynonymous mutations, including missence and nonsense. Missence mutations do not change amino acid but still have the possibility to influence the protein structure and

function, and eventually the trait expression. Nonsense mutation produces a premature stop/nonsense codon, usually resulting in a truncate, incomplete and nonfunctional protein product, and thus phenotype changes (Liu & Cordes 2004; Gjedrem & Baranski 2010; Yue 2014). Several approaches have been developed for SNP discovery and genotyping, and the most accurate and widely used method is DNA sequencing. However, this method is extremely expensive, time-consuming and low-throughput for discovery of a large number of SNPs and genotyping in a large population. Since the late 1990s, the development of microarray or gene chip-based genotyping platforms has dramatically increased the throughput of SNP genotyping in a large population, while greatly decreasing the cost per data point. Despite the greatly increased throughput of SNP arrays, up-front development cost is extremely high, which prevents the application of SNP arrays in most aquatic species (Liu & Cordes 2004; Gjedrem & Baranski 2010; Yue 2014).

Despite these achievements produced by the conventional markers, development of a large amount of markers based on Sanger sequencing and genotyping large populations are still slow, labor-intensive, time-consuming, and expensive. To robustly develop a large number of unbiased markers across the whole genome and high-density genotyping a large population at low cost, taking the advancement of next-generation sequencing (NGS), sequencing based marker discovery and genotyping was quickly adopted in aquatic species, which will be discussed in the following section.

1.4.2.2. Sequencing based marker discovery and genotyping

Inspired and driven by the quest of $1000 for a human genome, rapid advances of NGS technology have provided a great number of inexpensive data and revolutionized the genomics and breeding of plants and animals (Davey *et al.* 2011; Poland & Rife 2012; Andrews *et al.* 2016). This technology with far-reaching effect has greatly transformed the way genomes are sequenced, polymorphisms are discovered, populations are genotyped,

transcriptomes are sequenced and profiling is determined. Using this technology, it is possible to achieve high-throughput sequencing-based marker discovery and genotyping in a large population, and thus it has been quickly developed and adopted in many model and non-model organisms (Davey *et al.* 2011; Poland & Rife 2012; Andrews *et al.* 2016). The sequencing-based genotyping methods generally include multiplexing a proper number of samples with DNA barcodes, which are later used to separate each sample to reduce per sample cost. Only a small fraction of a genome is sequenced and yet the coverage is increased, improving the accuracy of genotyping in limited sequencing resources (Davey *et al.* 2011). A small portion of a genome can be archived by reducing the complexity of the genome by digestion with one or more restriction enzymes and the subsequent size selection due to the read-length limitation of current NGS platforms. These methods include, reduced-representation libraries (RRLs) and complexity reduction of polymorphic sequences (CRoPS), restriction-site-associated DNA sequencing (RAD-seq) and low coverage genotyping such as multiplexed shotgun genotyping (MSG) and genotyping-by-sequencing (GBS) (Davey *et al.* 2011). Among them, GBS is a widely used approach especially in non-model organisms without genome references or solid genomes (Elshire *et al.* 2011; Poland *et al.* 2012; Poland & Rife 2012). The original GBS employed one enzyme (Fig. 1.4). Later, an improved GBS approach has been developed, which employs two enzymes, one common cutter and one rare cutter. The combination of two enzyme cutters enables this approach to capture fragments associated with the rare cutters, which are roughly evenly distributed across the genome (Poland *et al.* 2012; Poland & Rife 2012). The size selection step filters fragment lengths from 300 to 800 bp and bridge-amplification further filters fragments only with one barcoded adaptor and one common adopter. The single end or both ends of hundreds of thousands fragments are then sequenced and millions of reads are produced, generating tens of thousands of unbiased SNPs spaced across the whole genomes. The feature of producing a larger amount of

unbiased markers in an inexpensive way, enables GBS to become the preferable approach to build high-density and high-resolution maps, facilitating QTL mapping and genomic selection, even map-based cloning (Poland & Rife 2012), which will be discussed later.

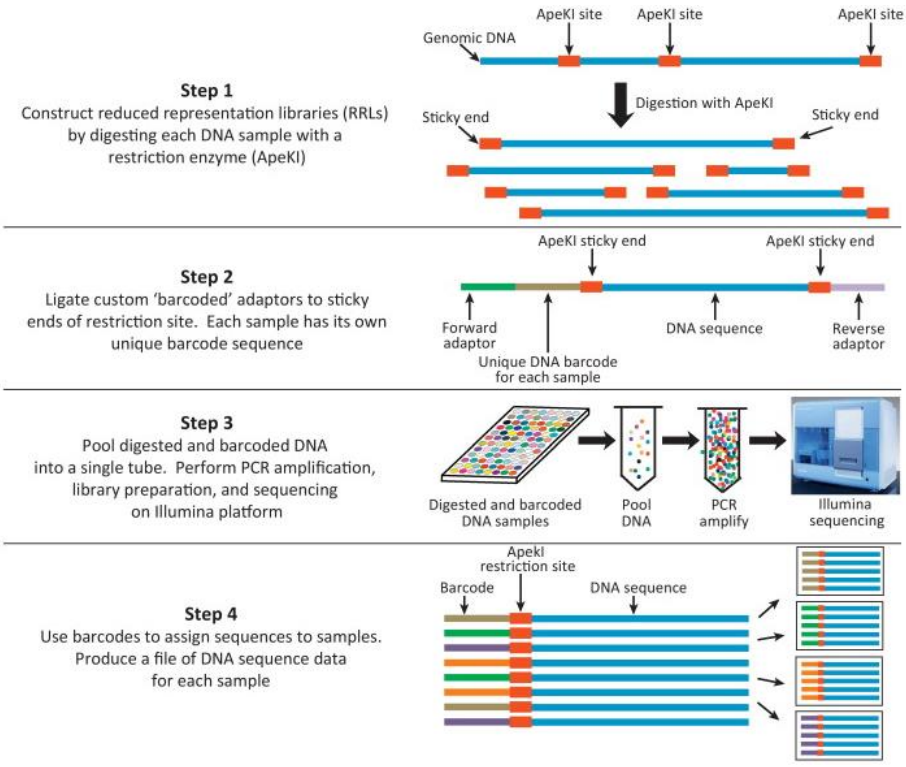


Fig. 1.4 Principle of genotyping-by-sequencing (GBS). Reprinted from Myles and his team (Myles 2013).

Riding the waves of sequencing-based marker discovery and genotyping, an increasing number of studies using this technology for various purposes have been reported in a variety of aquatic species in recent years. For example, thousands of markers were produced for investigation of genome analysis in rainbow trout (Sánchez *et al.* 2009; Hohenlohe *et al.* 2013) and cichlid species (Henning *et al.* 2014); growth related traits in clam (*Meretrix meretrix*) (Lu *et al.* 2013), hybrid tilapia (Liu *et al.* 2014a), large yellow croaker (*Larimichthys crocea*) (Ye *et al.* 2014), rainbow trout (Wang *et al.* 2015b) and Japanese flounder (*Paralichthys olivaceus*) (Cui *et al.* 2015); sex related traits in rainbow turbot (Vinas *et al.* 2012; Wang *et al.* 2015b) and Atlantic salmon (Gutierrez *et al.* 2014;

Ayllon *et al.* 2015; Barson *et al.* 2015); and other traits like swimming ability in common carp (*Cyprinus carpio*) (Laghari *et al.* 2014) and crowding in rainbow trout (Liu *et al.* 2015). Besides these important traits, it has also been used for disease related traits in several fish species, including common carp (Kongchum *et al.* 2010), and more recently, rainbow trout (Vallejo *et al.* 2014), Japanese flounder (Wang *et al.* 2014c) and channel catfish (*Ictalurus punctatus*) (Geng *et al.* 2015). However, to my best knowledge, no such study has been reported for disease related traits in Asian seabass.

1.4.3. Construction of linkage maps

1.4.3.1. Principle of linkage mapping

Linkage mapping is based on genetic techniques to generate maps showing sequence features including the position of genes and markers. A linkage map serves as a road map for identification of chromosomal regions and genes associated with traits of interest. Therefore, it is essential step for QTL mapping (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012). Markers or genes that are tightly linked or close together are more likely to co-segregate during meiosis, and thus are more likely to be transmitted together from parent to offspring than these that are located far away (Fig. 1.5). The recombination fractions can be determined by calculation of the proportion of recombinant gametes and are used to reflect the distances among markers. The larger the recombination fraction between markers, the further away they are located in the chromosome (Fig. 1.5). Markers with a recombination fraction of >50% between them are considered as unlinked (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012). Because the recombination fractions among markers are non-additive, mapping functions are used to convert recombination fractions into map units, which are referred to as centiMorgens (cM) (Fig. 1.6). Genetic distances measured by cM in a linkage map are linearly related and thus can be additive. Two widely-used mapping functions are Kosambi and Haldane. The

former assumes that the crossover events influence the adjacent crossover events, while the later assumes that there is no interference between recombination events (Kearsey & Harpal 1996; Harker *et al.* 2001). The likelihood of linkage between markers is determined by odds ratios of linkage versus no linkage. This ratio is more commonly presented as a logarithm of the ratio or logarithm of odds (LOD) value or score (Risch 1992). A LOD score of >3 between markers, which implies that the possibility of linkage (alternative hypothesis) is 1000 times more likely than the possibility of no linkage (null hypothesis), is generally considered as a critical value of linkage and thus used to construct a linkage map (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012).

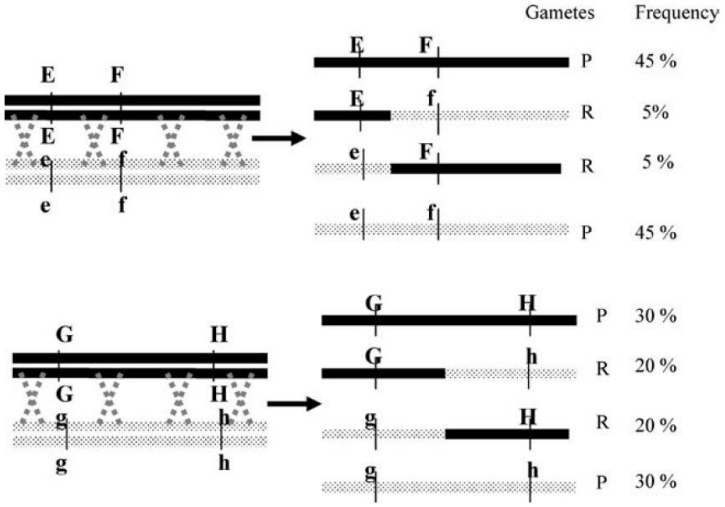


Fig. 1.5 Recombination events between homologous chromosomes during meiosis.
Gametes are either parental (P) or recombinant (R). The smaller the distance between two markers, the smaller the chance of recombination between them. Therefore, recombination between markers G and H should occur more frequently than recombination between markers E and F. This can be analyzed by the number of recombinants in a segregation population. Reprint from Collard and his team (Collard *et al.* 2005).

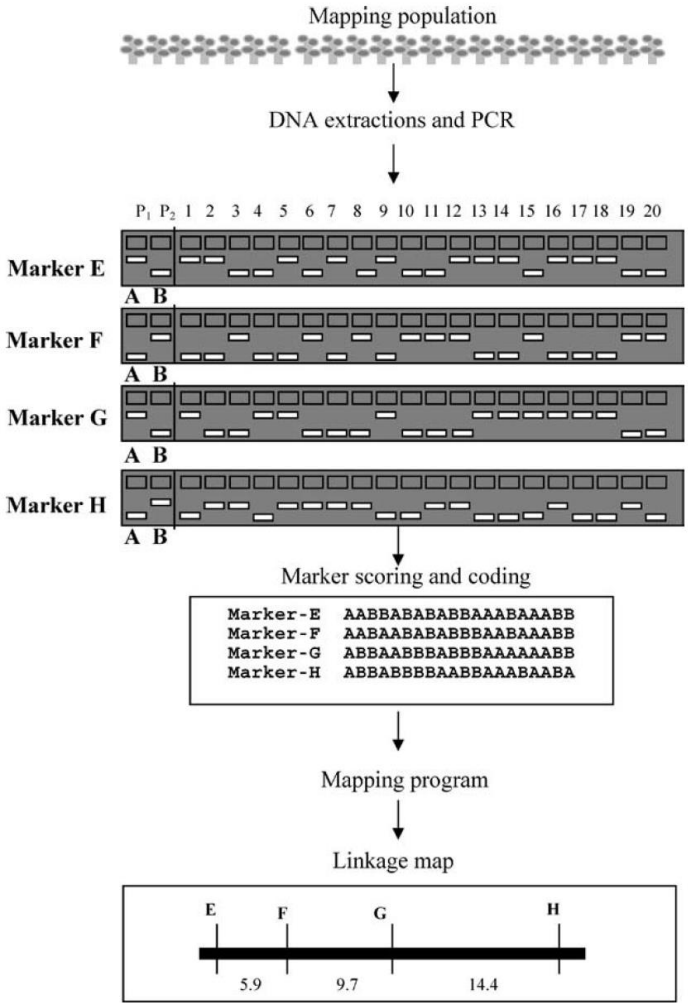


Fig. 1.6 Principle of construction of a linkage map. A and B represents parents P1 and P2, respectively. Reprint from Collard and his team (Collard *et al.* 2005).

1.4.3.2. Linkage maps in aquatic species

To date, using markers produced by conventional marker discovery and genotyping platform, linkage maps of over 45 fish species have been reported (Yue 2014). These species include Eastern (*Crassostrea virginica*) and Pacific (*Crassostrea gigas*) oyster, channel catfish, tilapia, Japanese flounder, yellowtail (*Seriola quinqueradiata*), common carp, zhikong scallop (*Chlamys farreri*), Atlantic salmon, rainbow trout, brown trout (*Salmo trutta*), black tiger shrimp (*Penaeus monodon*), kuruma prawn (*Marsupenaeus japonicus*), European seabass, blacklip abalone (*Haliotis rubra*), and Asian seabass (Gjedrem & Baranski 2010). However, these linkage maps constructed from conventional

markers have generally suffered from major disadvantages of low marker density, large marker interval and low resolution, as the number of markers in most of the linkage maps is less than 1000 and marker interval is more than 5 cM. Therefore, these linkage maps produce QTL with large confidence intervals, which could restrict their application in marker-assisted selection (MAS) programs (Gjedrem & Baranski 2010). Furthermore, precise QTL mapping and further determination of causative polymorphisms, even positional cloning of the corresponding genes, require a great number of genome-wide polymorphic markers and high-resolution linkage maps to saturate the linkage disequilibrium (LD) between marker and QTL (Goddard & Hayes 2009). It is very hard for such linkage maps to be constructed from conventional methods of marker development and genotyping to meet the increasing requirement for high density map and high resolution mapping (Davey *et al.* 2011).

Sequencing-based marker discovery and genotyping could overcome the difficulties of conventional methods to economically produce hundreds of thousands of SNPs in a large population. Therefore, high-density linkage maps, which could greatly enhance the high-resolution QTL mapping, could be expected in many aquatic species. For example, high-density maps were built for the orange-spotted grouper (*Epinephelus coioides*) with 4608 SNPs (You *et al.* 2013), Atlantic salmon with 6000 SNPs (Gonen *et al.* 2014), sea cucumber (*Apostichopus japonicas*) with 7839 markers (Tian *et al.* 2015), Pacific white shrimp (*Litopenaeus vannamei*) with 6359 markers (Yu *et al.* 2015), large yellow croaker with 3448 SNPs (Xiao *et al.* 2015), sticklebacks (*Gasterosteus aculeatus*) with 1001 and 978 markers (Glazer *et al.* 2015), Mexican tetra (*Astyanax mexicanus*) with 2200 markers (Carlson *et al.* 2015) and Asian seabass with 3321 SNPs (Wang *et al.* 2015a). These high-density maps have supplied resources for the accuracy of QTL mapping.

1.4.4. QTL mapping for disease resistance

1.4.4.1. Quantitative trait and QTL

Most economically important traits such as growth, flesh quality, feed conversion rate and disease resistance show continuous or quantitative variation, and are referred to as quantitative traits. They are inherited in a quantitative polygenic manner, and are generally controlled by a number of genes with small to moderate effects, and influenced by environment and gene-environment interactions (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012). The chromosome regions, harboring single gene or gene clusters and expressing the quantitative trait, are defined as QTL (Geldermann 1975). Identification of QTL or closely linked markers is extremely important as these markers have the potential to be applied in MAS to accelerate genetic gain for many traits. The reason is that these traits, such as disease resistance (causing morbidity and mortality), flesh quality (after death), longevity (late in life) and feed convention rate, are difficult and/or very expensive to measure (Gjedrem & Baranski 2010). There are two main approaches to determine QTL: candidate gene approach and QTL mapping approach. Candidate gene approach hypothesizes that a particular genetic variation in the gene has a large effect on the traits. Testing this hypothesis involves genotyping the gene in a proper number of individuals and evaluation of association between genotype and phenotype. The obvious disadvantage of this approach is the pre-requirement of information about of candidate gene, which has greatly limited its application (Gjedrem & Baranski 2010). The alternative approach, QTL mapping, could overcome this limitation.

1.4.4.2. Principle of QTL mapping

QTL mapping identifies genetic markers tightly linked to QTL which govern the quantitative traits. Unlike the candidate gene approach, QTL mapping assumes that the actual genes, which control the quantitative trait, are unknown. Instead, a proper number of genome-wide genetic markers are used to assess the association between genotypes of

the markers and phenotypes of traits. Based on the genotypes of markers, the mapping population is divided into different genotypic groups. If there is a statistically significant difference ($p < 0.05$) between the phenotypic means of genotypic groups, this implies that the marker could be linked to a QTL controlling the trait (Tanksley 1993; Young 1996). If there is a statistically significant difference ($p < 0.05$) between the phenotypic means of genotypic groups, this implies that the marker could be linked to a QTL controlling the trait (Tanksley 1993; Young 1996) (Fig. 1.7). The reason for a significant *p* value indicating a linkage between marker and QTL is due to recombination. Typically, the closer a marker and QTL are, the smaller the chance for crossover between the marker and QTL to occur. Vice versa: when there is a large distance between the marker and QTL, there is likely to be crossover events happening between them (Fig. 1.8). For the former case, the marker and QTL tend to be co-segregated during meiosis and inherited together in the offspring. Therefore, there will be significant difference of phenotypic means between genotypic groups of tightly-linked markers (Fig. 1.7). For the latter case, there will be no statistically significant difference of phenotypic means between genotypic groups for the loosely-linked markers (Fig. 1.8). In addition, no difference will be detected for the unlinked markers as these markers are located far away or in different chromosomes from QTL of interest and are randomly inherited with the QTL (Collard *et al*. 2005; Gjedrem & Baranski 2010; Boopathi 2012). The size and position of QTL can be determined through joint analysis of two or more markers from a linkage map.

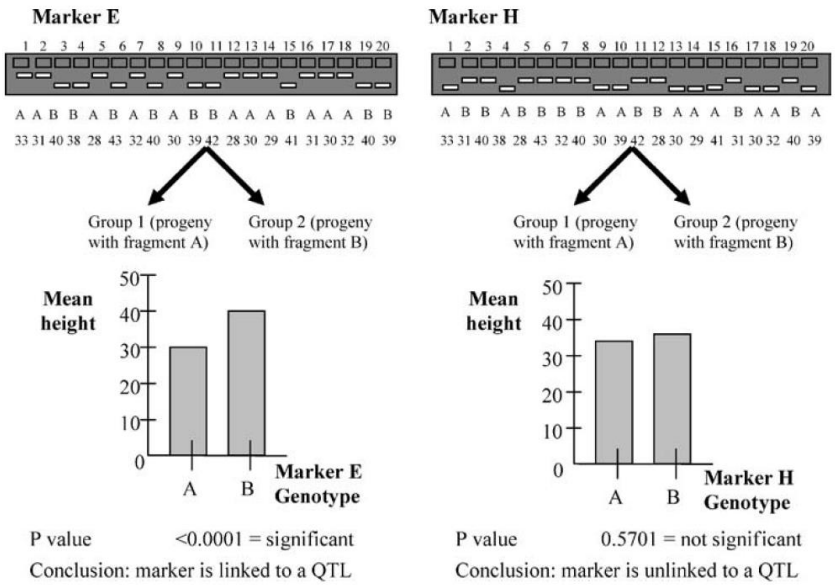


Fig. 1.7 Principle of QTL mapping. Markers that are linked to QTL or a gene controlling a particular trait will present significant differences between the partitioned populations based on the genotype of the marker. Marker E is linked to a QTL because of a significant difference between means, while marker H is unlinked to a QTL because of no significant difference between means. Reprint from Collard and his team (Collard *et al.* 2005).

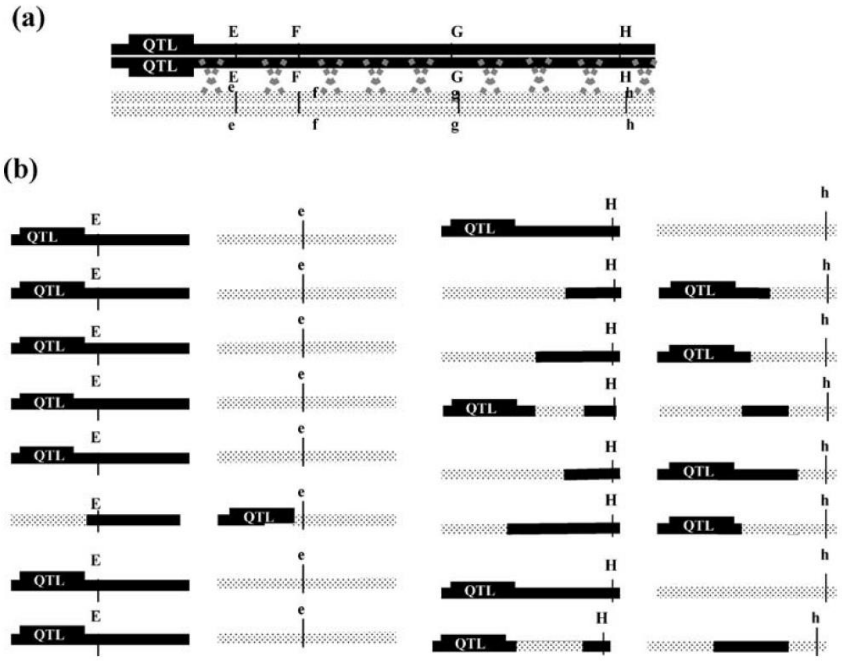


Fig. 1.8 Tight and loose linkage between markers and QTL. (a) Recombination events between QTL and markers. (b) Marker E tightly-linked to a QTL tends to inherit together with the QTL, while marker H loosely-linked to a QTL tends randomly to inherit with QTL. Reprint from Collard and his team (Collard *et al.* 2005).

Several methods have been developed to detect the QTL, including single-marker analysis, simple interval mapping (SIM) and composite interval mapping (CIM) (Tanksley 1993; Liu 1997). Among them, SIM is currently the most widely used one because it overcomes many disadvantages of variance analysis at marker loci (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012). This approach utilizes linkage maps and calculates confidence intervals between neighboring pairs of linked markers across the chromosome simultaneously. The analysis of linked markers could compensate for recombination between QTL and markers, and therefore more statistical power can be achieved (Lander & Botstein 1989; Liu 1997). The interval between adjacent pairs of markers is searched in an increment and whether QTL is present or absent within the interval is statistically tested. The SIM approach generates a profile of postulated sites for QTL between neighboring linked markers in a linkage map. The probability of a genotype (AA, Aa or aa), according to the marker genotype data, of an individual at the putative QTL is calculated. The highest probability of these genotypes is used to obtain the maximum likelihood (ML) and presented as LOD score. These LOD scores are used to determine the most likely position of QTL in the linkage map, where the largest LOD score is located. In addition, only these QTL with LOD scores exceeding a threshold value can be considered as statistically significant QTL (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012). The significance thresholds can be determined by permutation test. During the permutation test, each phenotypic value shuffles across the genotypic values of ordered marker which are kept constant. The shuffling results in a broken association between marker and trait, and the levels of false positives of marker and trait association are assessed. The process is iterated for a certain number of times (usually 1000 times). Based on the false positive level of marker and trait association, the significance level can be determined. The permuted 1000 LOD scores are compared with the observed LOD scores with the correct order of marker and trait value. The proportion of these permutated

1000 LOD scores that surpass the actually observed LOD score is presented as an approximate *p* value (Fig. 1.9) (Churchill & Doerge 1994; Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012).

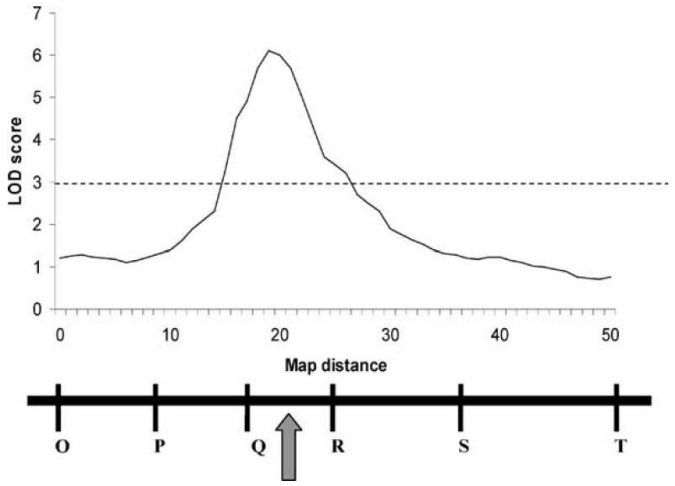


Fig. 1.9 Hypothetical output of a LOD profile in a linkage group. The dotted line represents the significance threshold determined by permutation tests. The output indicates that the most likely position for the QTL is near marker Q (indicated by an arrow). The best flanking markers for this QTL would be Q and R. Reprint from Collard and his team (Collard *et al.* 2005).

Family-based QTL mapping typically yields large confidence interval up to 20 cM, which means association between marker and QTL could be broken relatively rapidly through generations. This could decrease the precision of prediction using the linked markers and thus compromising the power of MAS. Increasing the linkage map's marker density and mapping population's size could reduce the confidence interval of QTL. This is referred to as fine mapping. Narrowing down QTL to a small region is necessary for efficient implementation of MAS. In addition, it is essential to identify potential candidate genes underlying the QTL, which could greatly enhance the understanding of the quantitative traits of interest (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012).

1.4.4.3. QTL mapping for disease resistance

Growth-related traits, meat quality, disease resistance and robustness are the most important traits in aquaculture. To accelerate the genetic gain and understand the genetic architecture of these quantitative traits, QTL mapping has been conducted in dozens of aquatic species, including Atlantic salmon, rainbow trout, European seabass, common carp and Asian seabass, as summarized in previous review papers (Gjedrem & Baranski 2010; Yue 2014).

Besides these important traits, QTL and association studies for disease resistance have also been reported in several fish species (Gjedrem & Baranski 2010; Ødegård *et al.* 2011a; Yue 2014). For example, twelve QTL for resistance to the parasite *Perkinsus marinu* have been mapped in the eastern oyster (Yu & Guo 2006). In addition, five QTL for resistance to *Bonamia ostreae* have been detected in the flat oyster (*Ostrea edulis*) (Lallias *et al.* 2009). Another five significant QTL for resistance to ostreid herpes virus type 1 (OsHV-1) have been identified in the Pacific oyster (Sauvage *et al.* 2010). For turbot, six QTL for resistance to *Aeromonas salmonicida* (Rodriguez-Ramilo et al. 2011), five for resistance to the parasite *Philasterides dicentrarchi* (Rodriguez-Ramilo *et al.* 2013), and several QTL for resistance to viral haemorrhagic septicaemia (VHS) (Rodriguez-Ramilo *et al.* 2014), have been identified. Furthermore, two QTL for resistance to IPNV (Ozaki *et al.* 2001), seven more QTL for resistance to IPNV (Ozaki *et al.* 2007), and 13 QTL for resistance to bacterial cold water disease (BCWD) (Wiens *et al.* 2013), have been mapped in rainbow trout. Additionally, 10 QTL for resistance to the parasite *Gyrodactylus salaris* (Gilbey *et al.* 2006), and one QTL for anaemia resistance (Moen *et al.* 2007), have been identified in Atlantic salmon. More recently, four QTL have been detected in Japanese flounder for resistance to *Vibrio anguillarum* infection (Wang *et al.* 2014c). These results are in consistent with the classical assumption that disease resistance is a complex and

polygenetic trait. Nonetheless, recently, this widely accepted assumption has been challenged by several studies. In 2006, a major QTL for resistance to lymphocystis disease (LD), explaining 50% phenotypic variation, was reported in Japanese flounder (Fuji *et al.* 2006). In 2008, Houston reported a major QTL for resistance to IPNV, with a phenotypic variation explained (PVE) of 24.6%, affecting the seawater stage of Atlantic salmon in Scotland (Houston *et al.* 2008). Surprisingly, in the following year of 2009, Moen independently reported the same single major QTL with a similar PVE of 23.4% for resistance to IPNV in the seawater stage of Atlantic salmon in Norway (Moen *et al.* 2009). Interestingly, in 2010, Houston again reported a major QTL with a PVE of 51%, affecting the susceptibility to IPNV in the freshwater stage of Atlantic salmon (Houston *et al.* 2010). Moreover, a major QTL for resistance to whirling disease, caused by the bacterium *Myxobolus cerebralis*, could explain 50–86% of the phenotypic variance across families in rainbow trout (Baerwald *et al.* 2011). These results could imply that disease resistance could be controlled by a few genes with large effect. Encouragingly, these major QTL have been implemented in MAS of Atlantic salmon (Ødegård *et al.* 2011a) and Japanese flounder (Fuji *et al.* 2007), leading to substantially reduced outbreaks of those diseases and economic losses. Despite these progresses in disease resistance, to the best of my knowledge, there is no such report in Asian seabass for resistance to VNN disease. Lacking such a study could severely undermine the sustainable production of Asian seabass.

Primary QTL mapping could improve our understanding about the genetic architecture about disease resistance. QTL fine mapping may help to identify possible candidate genes underlying QTL and thus improve our understanding about the mechanism of disease resistance. On the other hand, transcriptome analysis could provide an overview and nearly complete information about genes involved in host-pathogen interaction, and thus is another powerful tool which could help to identify disease resistance gene.

Furthermore, it could provide important information for identification of candidate gene and DNA marker in genes. For example, SSRs in genes, including UTRs, exons and introns, could be used as genetic markers for QTL mapping and association study, while gene expression data could supplement the identification of candidate genes underlying QTL.

1.5. Transcriptome analysis in aquaculture species

1.5.1. Principle of RNA-seq

The transcriptome is the complete set of transcripts and their quantities in a cell at a specific developmental stage or physiological condition (Wang *et al.* 2009). Investigation of the transcriptome profile is essential for functional analysis of the genome and understanding the molecular mechanism of development and disease (Wang *et al.* 2009; Ozsolak & Milos 2011; Mutz *et al.* 2013; Qian *et al.* 2014). Several technologies have been developed to deduce and quantify the transcriptome profile, such as hybridization-based microarray and sequence-based approaches. Microarrays are relatively high-throughput and inexpensive. However, it has been suffered from the need for prerequisite knowledge about gene sequences, high background noise due to cross-hybridization, limited capacity to detect a wide range of expression level and difficulty in comparing expression level across different experiments (Wang *et al.* 2009). Unlike microarrays, sequence-based approaches directly determine and quantify the cDNA sequences. These approaches include Sanger sequencing of cDNA or expressed sequence tag (EST) libraries, tag-based methods such as serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE) and massively parallel signature sequencing (MPSS). However, these Sanger sequencing-based methods are generally expensive, relatively low throughput and unable to completely sequence all the transcripts (Wang *et al.* 2009).

With the rapid development of NGS, this technology has been used to exploit the dynamic transcriptome, resulting in a technology termed as RNA sequencing (RNA-seq). RNA-seq has clear advantages over previous approaches and has revolutionized how the transcriptome is sequenced and gene expressions are quantified (Wang *et al.* 2009). It generally includes several steps (Fig. 1.10). Ribosome RNA (rRNA) are removed either through mRNA enrichment by oligo d (T) or depletion of rRNA by probe hybridization. Clean RNA are fragmented and converted into cDNA libraries. These libraries, either with or without PCR amplification, are then sequenced from single end (SE) or paired ends (PE) with a typical read length of 30-400 bp, depending on the sequencing platform. Hundreds of millions of short reads are then aligned to either a reference genome or transcriptome, or a *de novo* assembled transcriptome if there is no available reference genome. The structure and expression levels of transcripts are then determined (Wang *et al.* 2009). When designing the experiment of RNA-seq, several factors should be considered. 1) Gene expression is tissue specific, thus RNA samples should be collected from a proper number of tissues to reflect the addressed question. 2) Gene expression is time-dependent. A proper number of time points should be used to capture the whole biological process. 3) Biological replicates are essential for statistical analysis. Biological replicates could increase the statistical power, while reducing the false positives. 4) Minimum sequencing coverage must be met. A typical coverage of 30 times is enough for most gene expression analysis. For low and rare expression transcripts and a more complex transcriptome, higher coverage times are generally needed. 5) Sequencing read length (30-400) and type (SE or PE) are flexible, dependent on study purpose. For investigating expression levels of known genes in a species with a solid genome reference, short read length and SE sequencing with a low cost are generally acceptable. For discovering novel genes or gene expression in a species without a good reference, longer read length and PE with a higher cost are preferable (Wang *et al.* 2009; Ozsolak & Milos 2011; Mutz *et al.* 2013; Qian *et al.* 2014).

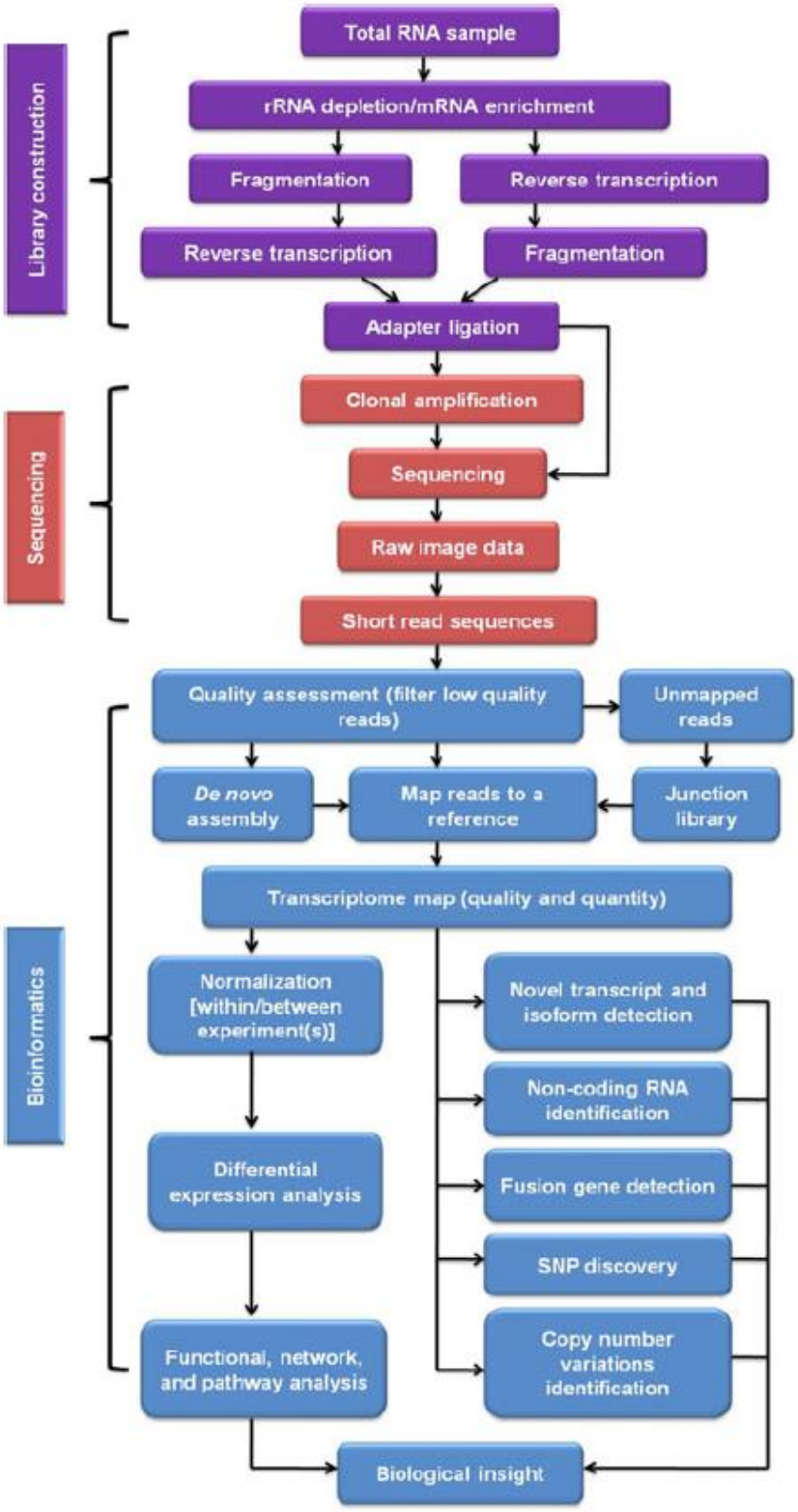


Fig. 1.10 Illustration of RNA-seq and data analysis pipeline. Reprint from Qian and his team (Qian *et al.* 2014).

1.5.2. RNA-seq for disease research in aquaculture species

Since its first launch in 2008 (Mortazavi *et al.* 2008; Nagalakshmi *et al.* 2008), RNA-seq has been widely used in various aspects of transcriptome related studies in a wide range of species. In aquaculture, RNA-seq has been quickly adopted in many aquaculture species for various purposes, including disease studies.

During the interaction between host and pathogen, fish are able to detect the presence of invading pathogens which triggers the first layer of defense through innate immune response (Ellis 2001). Successful defense against pathogen invasion relies heavily on the ability to rapidly activate and effectively mount strong innate immune responses, which leads to elimination of those pathogens in the host (Ellis 2001). Therefore, a very important strategy to reduce disease is to enhance innate immunity in the cultured fish species. Systematic investigation of molecular response such as transcriptome profiling, which could capture nearly complete information about host-pathogen interaction, is essential to understanding innate immunity. Since genome sequences of most aquaculture species are still not available, development of *de novo* transcriptome assembly algorithms, which could assemble millions of reads into large contigs or scaffolds, is critical (Ozsolak & Milos 2011). With the help of *de novo* assembly platforms such as Trinity (Grabherr *et al.* 2011; Haas *et al.* 2013), RNA-seq has the potential to capture nearly complete information about transcriptome dynamics during host-pathogen interaction. At the same time, it can simultaneously identify and quantify gene expressions with high sensitivity and accuracy at the genome level during the development of disease (Ozsolak & Milos 2011). Consequently, it could decipher the complicated gene regulation networks and signaling pathways activated and operated in disease development (Ozsolak & Milos 2011).

With its promising features, high throughput and declining operating cost, this technology has been applied to several aquatic species to investigate various diseases. For example, the transcriptome has been assembled and differentially expressed genes (DEGs) have been determined in Japanese seabass (*Lateolabrax japonicus*) infected with *Vibrio harveyi* (Xiang *et al.* 2010) and *V. anguillarum* (Zhao *et al.* 2016). In addition, transcriptome profilings of orange-spotted grouper in response to iridovirus (Huang *et al.* 2011) and NNV (Lu *et al.* 2012) infection have been determined. Similar work has also been reported in channel catfish infected with *Edwardsiella ictaluri* (Li *et al.* 2012). Furthermore, transcriptome profiling has been identified in Pacific white shrimp infected with taura syndrome virus (TSV) (Zeng *et al.* 2013). Similar works have also been reported in other species, such as blunt snout bream (*Megalobrama amblycephala*) challenged with *Aeromonas hydrophila* (Tran *et al.* 2015), half-smooth tongue sole (*Cynoglossus semilaevis*) during *V. anguillarum* infection (Zhang *et al.* 2015), grass carp (*Ctenopharyngodon idellus*) infected with *A. hydrophila* (Yang et al. 2016), and large yellow croaker in response to *Cryptocaryon irritans* (Wang *et al.* 2016). For Asian seabass, stress-responsive transcriptome and DEGs in the intestine have been determined (Xia *et al.* 2013b). Those transcriptome have facilitated the identification of DEGs in various pathways and biological processes, novel genes and studies of functional genomics. In addition, DNA markers, including SNPs, SSRs and insertion-deletion (indel), in those transcriptome have the potential to be applied in QTL mapping. Despite those progresses, transcriptome analysis and DEGs identification in response to NNV infection have not yet been reported in Asian seabass. Although rapid progresses have been achieved in identifying the transcriptome profiles of several fish species in response to a limited number of diseases, the large gap between comprehensive understanding of the mechanism of immune response and its applications in fish defense against disease still remains to be closed.

1.6. Rationales and objectives of this thesis

Despite the rapid expansion of the Asian seabass industry, VNN disease has remained as the major threat to its sustainable production. Mapping QTL for resistance to VNN disease and further identification of possible candidate genes are essential for understanding the genetic architecture of disease resistance and MAS in disease breeding program in Asian seabass. Furthermore, investigation of transcriptome profiling and DEGs in response to NNV infection is critical to understanding the molecular mechanism of Asian seabass-NNV interaction. However, to the best of my knowledge, such studies have not yet been reported in Asian seabass. Therefore, the overall purposes of the current study are to identify the genetic architecture of VNN disease resistance and transcriptome response to NNV infection to pave the way towards MAS for VNN resistant Asian seabass.

The specific objectives are:

1) To conduct primary QTL mapping for resistance to VNN disease in Asian seabass
2) To generate a large number of genome-wide SNPs using the NGS-based approach of GBS
3) To construct a high-density map
4) To conduct high-resolution QTL mapping
5) To identify candidate genes underlying these QTL
6) To sequence mRNA from NNV infected epithelial cells of Asian seabass using RNA-seq
7) To *de novo* assemble transcriptome of Asian seabass
8) To identify DEGs in response to NNV infection
9) To characterize top DEGs and their association with VNN disease resistance in Asian seabass

To achieve the abovementioned objectives, the following methods and procedures were used.

1) Artificially challenge Asian seabass fingerlings with NNV
2) Collect mortalities and surviving fish to form a panel
3) Genotype the panel with SSR markers for primary QTL mapping
4) Construct a linkage map using Joinmap 4.0 for primary QTL mapping
5) Conduct primary QTL mapping using MapQTL 6.0
6) Select 95 earliest mortalities and 95 surviving fish to form a second panel
7) Generate genome-wide SNPs and genotype the second panel using GBS
8) Construct a high-density map using Joinmap 4.0
9) Conduct a high-resolution QTL mapping using MapQTL 6.0
10) Challenge Asian seabass epithelial cells with NNV
11) Collect RNA samples from 6, 12, 24 and 48 hpi of mock and NNV challenged groups
12) Sequence the mRNA using Illumina HiSeq 2000
13) *De novo* assemble transcriptome of Asian seabass using Trinity platform
14) Identify DEGs by RSEM, edgeR and Blast2GO

The expected results of identification of QTL and the corresponding candidate genes will improve our understanding of the genetic architecture of QTL for disease resistance, and shed new light onto the mechanism of disease resistance in Asian seabass. In addition, the determination of markers closely linked to the corresponding QTL could possess the potential to be applied in MAS for disease resistance in Asian seabass, which is the ultimate goal for breeding. Determination of DEGs in response to NNV infection could help to understand the molecular mechanism of Asian seabass-NNV interaction and to identify disease resistance genes.

1.7. Project roadmap

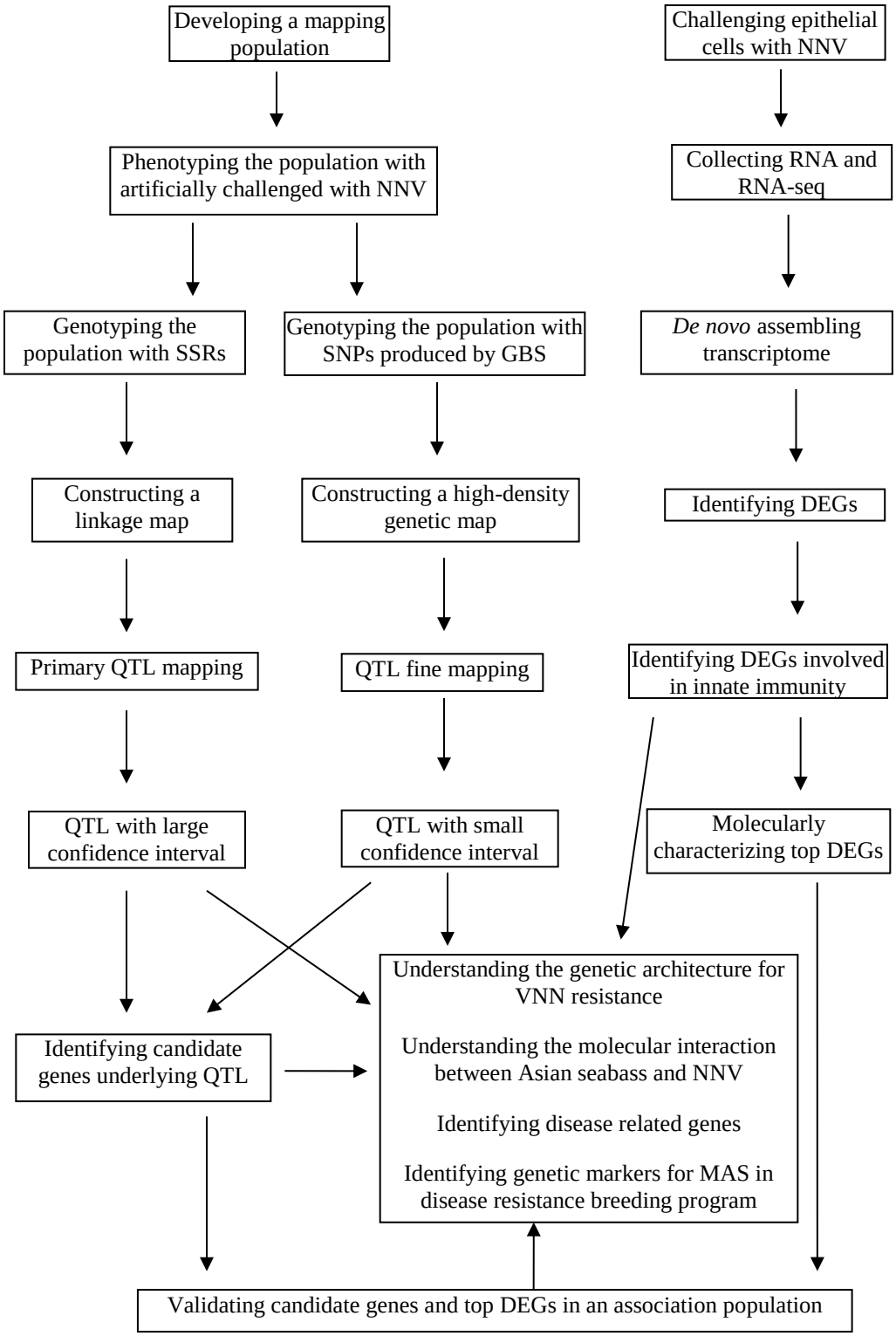


Fig. 1.11 Flowchart of experimental design.

Chapter 2. Primary QTL mapping for resistance to VNN disease in Asian seabass

2.1. Introduction

In mariculture, disease is one of the major restricting factors affecting the industry. Genetic improvement of disease resistance is an important task in this industry. In recent years, selective breeding programs for disease resistance have been initiated in several aquaculture species including Atlantic salmon (Thodesen & Gjedrem 2006; Houston *et al.* 2008), Atlantic cod (*Gadus morhua*) (Ødegård *et al.* 2010), rainbow trout (Dorson *et al.* 1995; Houston *et al.* 2010), rohu carp (*Labeo rohita*) (Nguyen & Ponzoni 2006) and Pacific white shrimp (Ødegård *et al.* 2011b). However, it is not trivial to improve disease resistance through conventional breeding due to difficulties in recording resistance traits in fish (Ødegård *et al.* 2011a). QTL mapping is an essential step towards MAS. Once markers associated with traits are determined, they could have the potential to be used in selection at fingerling stage (Ødegård *et al.* 2011a; Yue 2014).

QTL mapping for economically important traits has been widely conducted in domestic animals, helping to improve their performance substantially while generating huge economic benefits (Gjedrem & Baranski 2010; Gjedrem *et al.* 2012). On the other hand, the progress of QTL mapping on aquatic species is still behind that of domestic animals due to various reasons (e.g. too many species, less investment on each species). With the rapid advances of sequencing technologies and continually decreasing costs of genotyping, QTL and association studies for resistance against viral and bacterial diseases have also been reported in several fish species (Ødegård *et al.* 2011a; Yue 2014), and more recently in rainbow trout (Vallejo *et al.* 2014), Japanese flounder (Wang *et al.* 2014c) and

catfish (Geng *et al.* 2015). Major QTL for resistance against IPN and LD have been applied in MAS programs for Atlantic salmon (Gjedrem & Baranski 2010) and Japanese flounder (Fuji *et al.* 2007), leading to substantially reduced outbreaks of those diseases. Similar study has not yet been reported in Asian seabasss.

In the present study, Asian seabass fingerlings were artificially challenged with NNV and genotyped 520 (330 mortalities and 190 surviving) fingerlings with 145 microsatellite markers covering the genome of Asian seabass. Five significant and eight suggestive QTL for resistance, five significant and five suggestive QTL for survival time, were mapped against VNN in Asian seabass. QTL identified in this study could help understand the genetic architecture of VNN resistance in Asian seabass, and provide valuable information for fine mapping QTL for resistance to VNN disease.

2.2. Materials and methods

2.2.1. Ethics statement

All handling of fish in this thesis followed the instructions set up by the Institutional Animal Care and Use Committee (IACUC) of the Temasek Life Sciences Laboratory, Singapore, and the project was approved under the title "Breeding of Asian seabass resistance to viral diseases" (approval number TLL (F)-13-003) by TLL's IACUC.

2.2.2. Fish fingerlings used for challenge experiment

Fish used in this study were from a single backcross family (BC1 hybrid) of Asian seabass. Backcross populations are widely used for linkage mapping (Xu 2010). This approach is frequently employed to introduce a specific trait, such as disease resistance, to an elite line (Xu 2010). The sire used in this study, an offspring of the dam, was a survivor of an outbreak of big belly disease. Fingerlings at 30 dph were transferred from the Marine

Aquaculture Center (MAC), Agri-food and Veterinary Authority Singapore (AVA), located on St. John's Island, to Temasek Life Sciences Laboratory. The fingerlings were acclimated in a study tank with about 200 liters of circulated seawater (salinity 30 ppt, pH 7.6) at a stable temperature of 30°C, and saturated oxygen, for 7 days prior to challenge. Half of the seawater was replaced with fresh seawater every two days. The fingerlings were fed twice a day with a commercial diet (Marubeni Corporation, Tokyo, Japan) during the whole experiment. All the fish were closely monitored during the whole experiment.

2.2.3. Asian seabass epithelial cell line

The cell line was developed from the dissection of Asian seabass and was named SB cell line. It was originally obtained from a mixture of several cell types. However, only the epithelial-like cells survived after subsequent cultivation in the medium (Chong *et al.* 1987; Chang *et al.* 2001) and was provided by AVA. The SB cells were grown on Leibovitz's L-15 media (Life Technologies, Carlsbad, USA), supplemented with 10% fetal bovine serum (FBS) (Life Technologies, Carlsbad, USA) and 1% antibiotic antimycotic solution (Sigma-Aldrich, St. Louis, USA), in a 175 cm^2 flask (BD biosciences, Franklin lakes, USA) at 28°C. For subculture of SB cells, the cell monolayer was washed with 1x phosphate-buffered saline (PBS) with a volume equivalent to the volume of culture medium for 5 min. This process was repeated three times. To detach the SB cells from the bottom of the flask, about 3 ml of 0.1x trypsin (diluted in 1x PBS) (Sigma-Aldrich, St. Louis, USA) was added onto the washed cell monolayer in the 175 cm^2 flask. The flask was rotated several times to cover the cell monolayer with trypsin and subsequently incubated at room temperature for 5 min. The cells were examined under an inverted microscope to ensure that most cells were detached and floating in the medium. Any remaining attached cells could be released by gently tapping the side of the flask. The

released cells were resuspended in 3 ml of fresh medium and the trypsin was inactivated by FBS. Cell counting was performed in a hemocytometer using 200 µl of cell suspension. For cell culture maintenance, the cells of one flask were split into three flasks. For determination of virus titer, the cells were seeded into a 96-well plate at 5000 cells/well.

2.2.4. Preparation of NNV

The NNV strain GGNNV used in this study originated from the brain, head, kidney and liver of greasy grouper (*Epinephelus tauvina*) in Singapore (Chong *et al.* 1990). It belongs to RGNNV group, based on comparison of its genome sequence with the four members (Tan *et al.* 2001). It was also provided by AVA. The virus was propagated in SB cells. When 80% cell confluence was reached, NNV was added to the cell culture for propagation. The cells were examined using an inverted microscope every day. A typical cytopathic effect (CPE) caused by NNV was vacuolation in the infected SB cells. Once a full CPE was observed, cell culture was collected and kept at -80°C as virus stock. A total volume of about 2 L of virus suspension was obtained. About 500 ml of virus suspension was liquated into 500 cryo tubes (Thermo Fisher Scientific, Waltham, USA) and stored at -80°C. The cell culture containing the virus was freeze-thawed thrice before determination of virus titer. Virus titer of $TCID_{50}$ (TCID stands for tissue culture infective dose) was determined on SB cells in a 96-well cell culture plate with Karber's method (Kärber 1931).

2.2.5. Challenge of fingerlings with NNV

The immersion challenge method was adopted to provide data for genetic analysis based on a previously published paper (Huang *et al.* 2001; Fenner *et al.* 2006a). Compared with the injection approach, immersion challenge could represent a more natural route for pathogen infection (McBeath *et al.* 2015). In addition, it could challenge hundreds and

thousands of fish simultaneously within a controlled time, greatly increasing the challenge throughput. On the day of challenge, 18 fingerlings were randomly selected as a subgroup for assessing the average body weight of the whole population. And it was estimated to be 1.00 ± 0.20 g. Subsequently, 700 fingerlings were transferred to a 30-liter tank with clean seawater containing 9×10^6 $TCID_{50}$/ml of NNV and immersed for two hours under close monitoring. For the control, similar to the NNV challenge tank, 700 fingerlings were sham challenged in clean seawater by adding a volume of used L-15 media equivalent to that of the NNV containing cell culture used in the challenge tank. After that, each group of 700 fingerlings was transferred back to a tank with clean seawater. All the challenge experiments described in this thesis followed this procedure.

2.2.6. Measurement of traits

The behavior of the fingerlings was closely observed during the whole experiment. Twice a day, mortalities were recorded and removed, stored in pure ethanol and kept at -80°C for further analysis. The whole experiment was terminated when there were less than three mortalities for two consecutive days. All the survivors from the mock and 200 survivors among the NNV challenged fingerlings were killed by overdose of AQUI-S® and stored in pure ethanol at -80°C. Two disease-related traits, resistance and survival time, were recorded. Resistance was a dichotomous trait with binary data defined as survival (trait value of 1) vs. non-survival (trait value of 0) status of each individual at the end of the experiment. Thus, this model excluded time of survival from trait analysis. Such data can be analyzed using a linear model and a precise prediction of breeding value can be achieved (Ødegård *et al.* 2011a; Rodriguez-Ramilo *et al.* 2011; Rodriguez-Ramilo *et al.* 2013; Rodriguez-Ramilo *et al.* 2014). Resistance was based on whether each challenged fish was dead/alive at a certain time point and ignored the lifespan of the fish. On the other

hand, to some extent, survival time may also reflect the degree of resistance. For example, the fish that died later might have been more resistant than the fish that died earlier. Thus, survival time was also recorded. It was defined as the number of days post challenge (dpc) for which each individual lived (e.g. give trait value of 10 for fish that died at 10 dpc). Consequently, the survival time for all the survivors was scored with the same value (e.g. give trait value of 25 for all the surviving fish). A survival score model has been proposed to generate the data (Veerkamp *et al.* 2001), which can be analyzed using a threshold model with a more accurate prediction of breeding value in some cases (Ødegård *et al.* 2011a; Rodriguez-Ramilo *et al.* 2011; Rodriguez-Ramilo *et al.* 2013; Rodriguez-Ramilo *et al.* 2014). All the trait measurements described in this thesis followed this procedure.

2.2.7. RNA isolation from the brains of challenged fingerlings

The NNV was examined on mortalities and survivors from both control and challenged groups. From each group, five fingerlings were randomly selected from pools of mortalities and survivors, and subsequently dissected to collect the brain tissues. Total RNA was extracted from the collected brains using the TRIzol reagent (Life Technologies, Carlsbad, USA) following the manufacturer's instructions. In brief, a brain sample was homogenized in 200 µl TRIzol using a power homogenizer before adding 800 µl more TRIzol to a final volume of 1 ml. The homogenized sample was incubated at room temperature for 5 min to ensure complete dissociation of nucleoprotein complexes. After that, 200 µl of chloroform was added to the lysate and mixed well with the sample through vortexing before incubating at room temperature for 3 min. The mixture was centrifuged at 12000 *x* g for 15 min at 4°C. Following the centrifugation, 450 µl of the supernatant was transferred to a fresh tube before adding 500 µl of ice-cold isopropanol. The mixture was incubated at -20°C for 1 h before centrifugation at 12000 *x* g for 20 min at 4°C to precipitate

RNA. The supernatant was completely removed and the RNA pellet was washed by adding 1 ml of ice-code 75% ethanol before centrifugation at 8000 *x* g for 10 min at 4°C. Following the RNA washing, the ethanol was completely removed and the RNA pellet was dried at room temperature for 15 min before adding 20 µl of nuclease-free water. The RNA was incubated at 65°C for 10 min to ensure the complete dissolution of RNA pellet. To assess RNA quality, 1 µl of RNA was electrophoresed on 1% agarose gel stained with 1% ethidium bromide (EB) and visualized using Gel DOC XR (Bio-Rad Laboratories, Hercules, USA). RNA concentration was determined by Nanodrop (Thermo Fisher Scientific, Waltham, USA). The RNA sample was kept in -80°C for downstream analysis. All RNA extraction described in this thesis followed this procedure.

2.2.8. Converting RNA to cDNA by reverse-transcription

Before reverse-transcription, DNA contamination in the RNA samples was removed using DNase I (Roche, Basel, Switzerland) following the manufacturer's instructions. In brief, 25 µl of reaction volume contained 2.5 µl of 10 x incubation buffers, 2 units of DNase, 2 units of RNase Inhibitor (Promega, Madison, USA), 4 µg of RNA and RNase-free water. The samples were incubated at 37°C for 20 min to digest the genomic DNA and finally incubated at 75°C for 20 min to inactivate DNase. Following digestion of genomic DNA, reverse-transcription to convert RNA to cDNA was performed using M-MLV Reverse Transcriptase (Promega, Madison, USA) following the manufacturer's instructions. In brief, 2 µg of the digested RNA and 0.5 µg of random hexamer oligo were mixed. The mixture was heated at 70°C for 5 min before immediately cooling on ice to prevent formation of secondary structure. Reverse-transcription was conducted in a reaction volume of 40 µl containing 15 µl of mixture of RNA and random primers, 5 µl M-MLV 5 x reaction buffer, 1.25 mM dNTPs, 25 units of RNase Inhibitor, 200 units of M-

MLV Reverse Transcriptase and Nuclease-free water. The mixture was incubated at 37°C for 60 min to form the first-strand of cDNA. The cDNA samples were diluted 5x with Nuclease-free water and kept at -80°C for downstream analysis. All the cDNA synthesis in this thesis followed this procedure.

2.2.9. Examination of NNV

Two pairs of primers (RNA1-4 and RNA2-4, Appendix Table A1) specifically and respectively targeting RNA1 and RNA2 of NNV were designed using PrimerSelect in DNASTAR Lasergene version 11 (Lasergene, Madison, USA), and subsequently used in multiplex PCR amplification to detect the presence of NNV in the examined samples. The expected lengths of PCR products were 826 and 426 bp, respectively. A 25 µl of PCR reaction included 1 x PCR buffer, 500 µM of each dNTP, 1.5 µM of primers of RNA1-4F and RNA1-4R, 2.5 µM of primers of RNA2-4F and RNA2-4R, 2.5 units of Taq polymerase and 8 ng of cDNA. The thermal cycling conditions were as follows: 94°C for 3 min, followed by 36 cycles of 94°C for 30 s, 59°C for 30 s and 72°C for 45 s, with a final extension at 72°C for 10 min. Water was used as a negative control and cDNA from NNV was used as positive control. *α-Tubulin* (Srichanun *et al.* 2013) was also amplified as an assessment of the RNA integrity using a pair of primers of Lca-tub shown in Appendix Table A1 (5 µM of each in 25 µl PCR reaction) and annealing temperature of 55°C. PCR product of 5 µl was electrophoresed on 2% agarose gel stained with 1% EB and DNA fragments were visualized using Gel DOC XR (Bio-Rad Laboratories, Hercules, USA). All the primer pairs used in this thesis were designed using PrimerSelect in DNASTAR Lasergene. All gel electrophoresis described in this thesis followed this procedure.

2.2.10. DNA isolation

A total of 330 out of 347 mortalities and 190 randomly selected survivors from the challenged group were used as a panel for data collection. Caudal fins of the panel were taken. DNA was isolated using a method developed by our lab (Yue & Orban 2005). In brief, about 0.16 mm^2 of fin tissue was cut and ethanol on it was removed. The fin cutting was homogenized in the extraction buffer containing protenase K (final concentration of 400 µg/ml) (Sigma-Aldrich, Missouri, USA) at 55°C and 160 *x* g for two hours. Sample lysate was centrifuged at 4000 *x* g for 5 min. To activate the cilica, 90 µl (6 M) NaI was added to a 96-well Pall filter column plate (Pall Corporation, Washington, USA). The plate was incubated at room temperature for 3 min before adding 30 µl of the lysate supernatant. The mixture was incubated at room temperature for 2 min before centrifugation at 4000 *x* g for 2 min. DNA remaining in the column was washed with 240 µl washing buffer by centrifugation at 4000 *x* g for 2 min. Ethanol was evaporated by incubating the plate at room temperature for 5 min. To elute DNA, 120 µl water was added to the column and incubated at room temperature for 15 min before collection by centrifugation at 4000 *x* g for 5 min. DNA concentration was determined by Nanodrop (Thermo Fisher Scientific, Waltham, USA). To assess DNA quality, 2 µl of DNA was electrophoresed on 2% agarose gel stained with 1% EB and visualized using Gel DOC XR (Bio-Rad Laboratories, Hercules, USA). A proper amount of DNA was diluted to a final concentration of 10 ng/µl as working stock and kept in -80°C for downstream analysis.

2.2.11. Microsatellites markers and population genotyping

A total of 148 microsatellite markers were selected based on a published linkage map (Wang *et al.* 2011). In addition, one SSR marker LcaRTP3 was recently developed from gene receptor-transporting protein 3 (*rtp3*) in Chapter 4 and used in this study. These markers, roughly present in even distribution in the genome and achieving a wide coverage

within and across all the 24 linkage groups, were used to genotype the panel. To achieve a relatively higher throughput of genotyping, a fluorescence-based genotyping method was used. Fluorescence bases were labeled to primers and subsequently incorporated into PCR products during PCR reaction. The lengths of PCR products with fluorescence bases can be determined using capillary electrophoresis. Either forward or reverse primers were labeled with either FAM or HAX fluorescence. PCRs were conducted as described in (Wang *et al.* 2011). In brief, 25 µl of reaction volume contained 1 x PCR buffer, 500 µM of each dNTP, 0.2 µM of each primer, 2.5 units of Taq polymerase and 10 ng of DNA. The PCR reactions were performed on the thermal cycling with the following conditions: 94°C for 3 min, followed by 36 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 45 s, with a final extension at 72°C for 10 min. PCR product of 5 µl was electrophoresed on 2% agarose gel stained with 1% EB, DNA fragments were visualized using Gel DOC XR (Bio-Red, Hercules, USA) and their amounts were assessed according to the brightness of the bands. Typically, 10 to 20 PCR products were multiplexed to reduce genotyping cost and increase throughput, and diluted an appropriate number of times before submitting to the DNA analyzer. PCR products were analyzed using capillary electrophoresis by a 3730xl DNA analyzer (Applied Biosystems, California, USA) and allele sizes were determined by comparison with size standard GS-ROX-500 (Applied Biosystems, California, USA) using the software Genemapper (Applied Biosystems, California, USA). All PCR and genotyping of SSRs described in this thesis followed this procedure.

2.2.12. Linkage mapping

Linkage mapping is based on genetic techniques to generate maps showing sequence features including the positions of genes and markers. A linkage map serves as a road map for identification of chromosomal regions and genes associated with traits of

interest. Therefore, it is an essential step for QTL mapping (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012). In this study, a total of 149 microsatellite markers, were used for linkage mapping. The criteria for marker selection were that the number of markers per linkage group (LG) would be evenly distributed as much as possible and the average distance between markers should not be too large, with recommended 20 cM for QTL detection (Dekkers & Hospital 2002). The assessment for Mendelian segregation distortion of each marker using Chi-square test was performed on the software JoinMap 4.1 (Van Ooijen 2011). The score of deviation from the expected ratio 1:1 of each marker was checked and the marker was removed from further analysis if it was deemed unreliable for determining genotype. Linkage among markers was evaluated using JoinMap 4.1 with mapping function of Kosambi. Marker order and position in the linkage map were determined if a minimum LOD score exceeded three as an indication of significant linkage (Van Ooijen 2011). The linkage map was visualized using Mapchart (Version 2.2) (Voorrips 2002). All linkage maps described in this thesis were constructed using JoinMap 4.1 with the same parameters.

2.2.13. QTL mapping

The main function of construction of a linkage map is to identify chromosome regions and/or genes that are responsible for the trait of interest by determining the association between phenotype and genotype of markers through QTL mapping. After linkage analysis, identification and mapping of QTL were carried out by MapQTL 6 (Van Ooijen & Kyazma 2009). The confidence intervals were estimated by bootstrapping methods to define the smallest chromosome segment with 95% of the most likely QTL position (Visscher *et al.* 1996). The association between marker and QTL was determined by the Kruskal-Wallis analysis (Van Ooijen & Kyazma 2009). To determine the statistical

significance of the QTL signal, the significant threshold of LOD was determined through a simulation with a permutation test of 1000 times for each LG and trait under the null hypothesis of no QTL at a given map position (Doerge & Churchill 1996). QTL with LOD scores greater than threshold scores at $p < 0.05$ level at chromosome-wide were considered as suggestive, while greater than threshold scores at $p < 0.05$ and $p < 0.01$ at genome-wide were considered as significant and very significant, respectively. All QTL mapping described in this thesis were constructed using MapQTL 6 with the same parameters.

2.3. Results and discussion

2.3.1. Challenge experiment and trait measurement

To collect phenotypic data of levels of disease resistance, a back-cross mapping population consisting of 700 Asian seabass fingerlings was generated and challenged at 37 dph with NNV at a concentration of 9×10^6 $TCID_{50}$/ml for two hours. All the fish were closely monitored and daily mortalities were recorded and collected. Clinical signs such as dark coloration, lying down on the bottom, abnormal swimming patterns of 'whirling' and head down were observed in most challenged fish (Fig. 2.1 b and c). These are typical symptoms of NNV infection (Shetty *et al.* 2012) and are consistent with the symptoms previously reported in infected Asian seabass artificially challenged with NNV (Huang *et al.* 2001). The symptoms imply that the challenged Asian seabass had successfully been infected by NNV. In contrast, in the control group, dark coloration was observed in only a few of the fish and abnormal swimming behaviors were not observed (Fig. 2.1 a and c). As shown in Fig. 2.2, mass mortality (more than 4 a day) was observed in the NNV-challenged tank after 10 dpc with a peak of 37 mortalities a day at 20 dpc. A similar result was also reported in a previous study, in which the mass mortality began at 9 dpc in the NNV challenged Asian seabass (Fenner *et al.* 2006a). In addition, this result is also in line with

that of a study on IPNV-challenged Atlantic salmon, in which mass mortality started at 10 dpc with a peak at 23 dpc (Houston *et al.* 2010). For the sham-challenged control, a slight rising mortality also occurred in the control tank at approximately the same time (20 dpc), with unclear reasons. Interestingly, in the above-mentioned study, there was also a small rise of mortality in the control tank of Atlantic salmon with unclear reasons (Houston *et al.* 2010). The mortalities decreased to close to baseline (less than three mortalities for two consecutive days) at 24 dpc. Hence, the experiment was terminated at that time. The overall mortality rates of the mock group and challenged group were 15.29% and 49.57%, respectively. Termination of experiment when cumulative mortality rate approaches about 50% is a typical practice in challenge experiments (Fjalestad *et al.* 1993; Ødegård *et al.* 2011a).

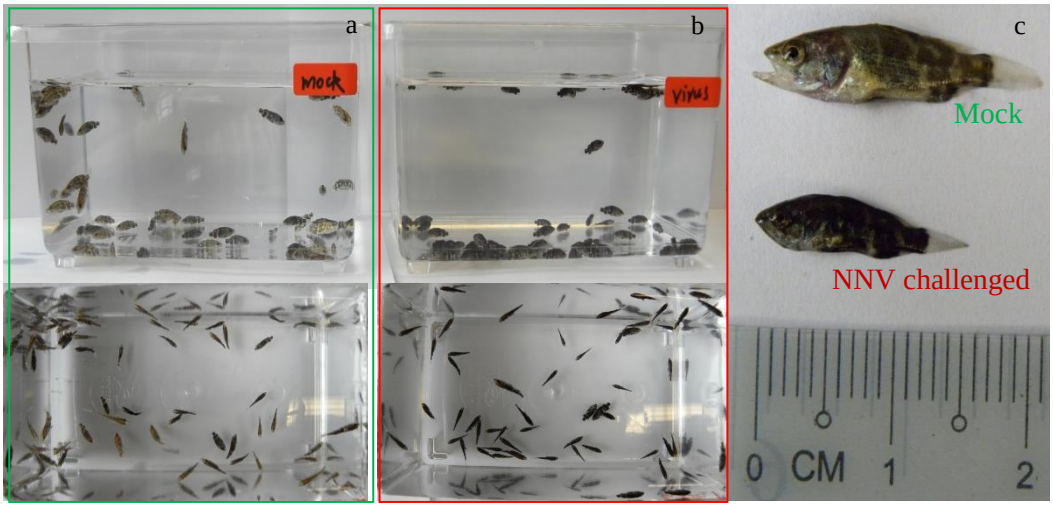


Fig. 2.1 Mock and VNN-challenged Asian seabass fingerlings. First and second rows of a and b are the front and top views, respectively. a shows mock fish, displaying more activity and less dark coloration in their bodies. b shows NNV-challenged fish, displaying clinical signs of less activity (floating and resting at the bottom of tank) and dark coloration in their bodies. c shows a mock fish (upper) and a NNV-changed fish (lower). Dark coloration was observed in the NNV-challenged fish but not in the mock.

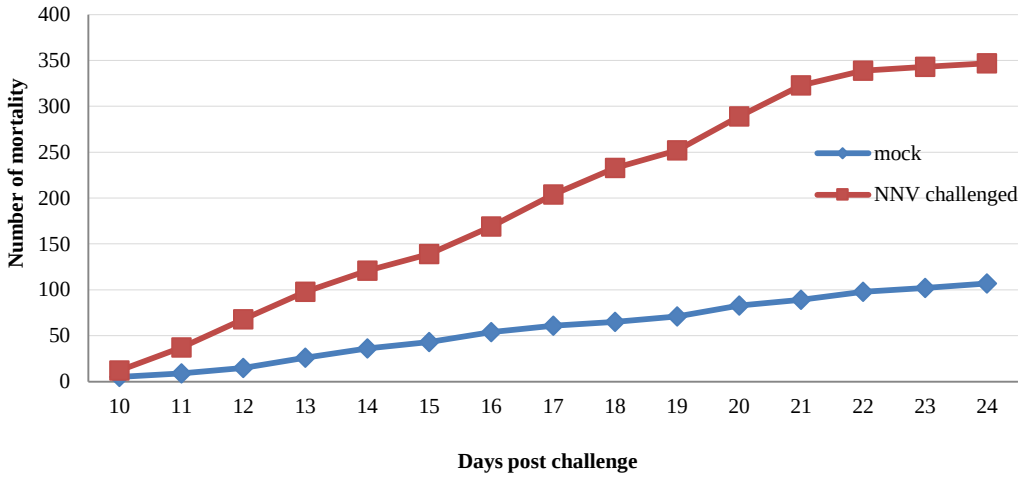


Fig. 2.2 Cumulative mortality curves of mock and NNV-challenged Asian seabass fingerlings.

2.3.2. Examination of NNV

To confirm the successful infection by NNV and that the subsequent mortality in the challenged group was caused by NNV infection, molecular detection of the NNV genome by RT-PCR was performed in the challenged and control groups. Two primer pairs were designed to target RNA1 and RNA2 of NNV, respectively. RNA extracted from brains, which are the main target organ of NNV, of both mortalities and survivors of the challenged and control fish, were converted to cDNA. PCR amplification on the cDNA showed that the two expected bands (826 bp for RNA1 and 426 bp for RNA2) were detected, from survivors (Fig. 2.3, lanes 11-15) and mortalities (Fig. 2.3, lanes 16-20) in the NNV-challenged group. Further analysis showed that bands detected in mortalities were more intensive than those from survivor samples. There was no amplification from survivors (Fig. 2.3, lanes 1-5) and mortalities (Fig. 2.3, lanes 6-10) in the mock group. It is worth noting that a detectable amount of NNV was present in the surviving fish from the challenged group although they did not show any symptoms of VNN. This asymptomatic state seems very common in aquatic species and could pose great a threat to the population

via both horizontal (co-habitation) and vertical transmission of disease (Nakai *et al.* 2009; Gjedrem & Baranski 2010; Shetty *et al.* 2012). It has long been recorded in several fish species including Asian seabass (Azad *et al.* 2006), European seabass (Hodneland *et al.* 2011; Haddad-Boubaker *et al.* 2013), seabream (Castric *et al.* 2001; López-Muñoz *et al.* 2012; Haddad-Boubaker *et al.* 2013), Atlantic cod (Rise *et al.* 2010), turbot (Olveira *et al.* 2013), and grouper (Mao *et al.* 2013), in both farmed and wild populations. This phenomenon has two implications. On one hand, those fish are the carriers of the virus and could be the potential sources for horizontal and vertical transmission of VNN. In addition, these asymptomatic fish could pose more threats than the acute infected ones, as diseased fish could be removed timely to prevent disease transmission, whereas asymptomatic fish are hard to identify, and thus could not be removed. Therefore, development of a feasible and efficient method to detect the virus and separate asymptomatic individuals deserves further investigation. On the other hand, it is suspected that the genes in the mapped QTL (discussed later) in this study may not be involved in host resistance in terms of preventing virus infection, but could be involved in pathways for limiting virus replication. Hence, more investigation in those genes will be carried out.

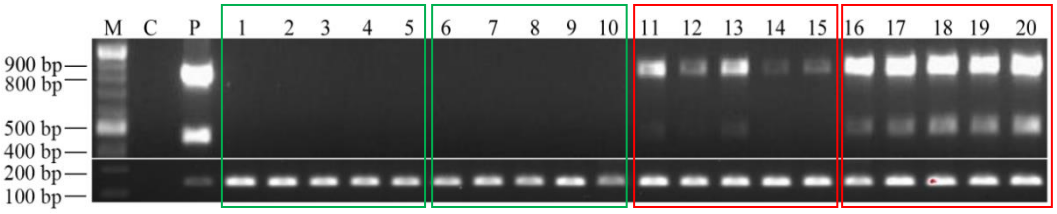


Fig. 2.3 PCR detection of NNV in the brains of sampled Asian seabass fingerlings. Upper panel: fragments amplified by RNA1-4 and RNA2-4. Lower panel: *α-Tubulin*. M: marker. C: negative control. P: positive control. Lane 1-5: survival fingerlings (mock). Lane 6-10: mortalities (mock). Lane 11-15: survival fingerlings (NNV-challenged). Lane 16-20: mortalities (NNV-challenged).

2.3.3. Linkage mapping

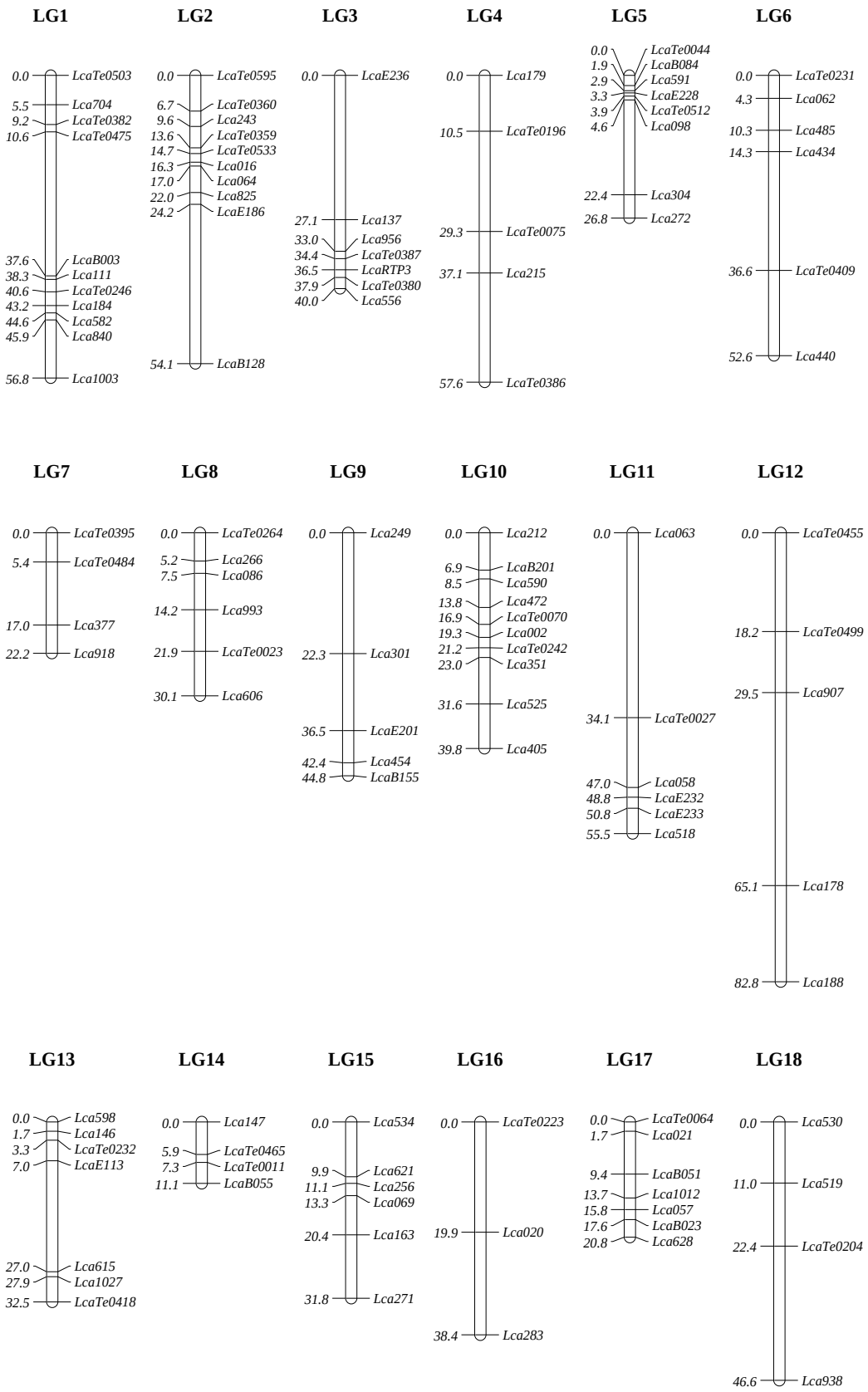
Linkage mapping is based on genetic techniques to generate maps showing sequence features including the position of genes and markers, based on the recombination events during the meiosis. A genetic/linkage map servers as a road map for identification of chromosomal regions and genes associated with traits of interest. Therefore, it is essential for QTL mapping (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012). In this study, a total of 148 microsatellites markers roughly evenly distributed in the 24 LGs from the second generation linkage map of Asian seabass were selected (Wang *et al.* 2011). An additional SSR marker from the 3'UTR of *RTP3* was also selected. Those markers were used to genotype the mapping population of 522 individuals. The size of the mapping population for a preliminary linkage mapping generally ranges from 50 to 250 individuals (Mohan *et al.* 1997). The population size is also influenced by the genome size and complexity of the species being investigated, as a larger and more complex genome generally requires a larger mapping population size. In addition, a larger population is required for high-resolution mapping as it could have more recombinants (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012). Previous studies of QTL mapping for disease resistance in turbot used a population of 270 individuals for VHS (Rodriguez-Ramilo *et al.* 2014), and 400 individuals for furunculosis (Rodriguez-Ramilo *et al.* 2011). In this study, the mapping population size is much larger and could yield a better mapping resolution. Among the 149 markers, 145 were successfully mapped on the linkage map consisting of 24 LGs spanning 994.06 cM (Table 2.1, Fig. 2.4), which is much shorter than a previous linkage map of 2411.5 cM containing 790 markers (Wang *et al.* 2011). The reason could be due to the different families, and smaller number of markers used in this study, resulting in some missing representations of regions in the genome. The number of markers in each LG varied from 3 to 11 with an average number of 6.04 markers per LG (Table 2.1, Fig. 2.4). The lengths of each LG varied from 11.09 to 82.76 cM with an average length of 41.42 cM. The average distance between markers was 8.22 cM , which

is enough for a primary QTL mapping as the recommended minimum marker interval is 20 cM (Dekkers & Hospital 2002). Nonetheless, further studies could focus on generating more markers and constructing a high-density map to improve the power of QTL detection. This will be discussed in Chapter 3.

Table 2.1 Statistics of the re-constructed linkage map of Asian seabass.

LG	N. of markers	Total length (cM)	Marker interval (cM)
1	11	56.81	5.68
2	10	54.13	6.01
3	7	40.03	6.67
4	5	57.56	14.39
5	8	26.77	3.82
6	6	52.62	10.52
7	4	22.23	7.41
8	6	30.12	6.02
9	5	44.76	11.19
10	10	39.78	4.42
11	6	55.51	11.10
12	5	82.76	20.69
13	7	32.52	5.42
14	4	11.09	3.70
15	6	31.82	6.36
16	3	38.38	19.19
17	7	20.75	3.46
18	4	46.63	15.54
19	6	41.06	8.21
20	6	38.72	7.74
21	5	58.37	14.59
22	3	29.93	14.97
23	5	31.99	8.00
24	6	49.75	9.95
total	145	994.06	-
minimum	3	11.09	3.46
maximum	11	82.76	20.69
average	6.04	41.42	8.21

Note: LG: linkage group; N.: number; -: not available


LG1
0.0 LcaTe0503
5.5 Lca704
9.2 LcaTe0382
10.6 LcaTe0475
37.6 LcaB003
38.3 Lca111
40.6 LcaTe0246
43.2 Lca184
44.6 Lca582
45.9 Lca840
56.8 Lca1003
LG2
0.0 LcaTe0595
6.7 LcaTe0360
9.6 Lca243
13.6 LcaTe0359
14.7 LcaTe0533
16.3 Lca016
17.0 Lca064
22.0 Lca825
24.2 LcaE186
54.1 LcaB128
LG3
0.0 LcaE236
27.1 Lca137
33.0 Lca956
34.4 LcaTe0387
36.5 LcaRTP3
37.9 LcaTe0380
40.0 Lca556
LG4
0.0 Lca179
10.5 LcaTe0196
29.3 LcaTe0075
37.1 Lca215
57.6 LcaTe0386
LG5
0.0 LcaTe0044
1.9 LcaB084
2.9 Lca591
3.3 LcaE228
3.9 LcaTe0512
4.6 Lca098
22.4 Lca304
26.8 Lca272
LG6
0.0 LcaTe0231
4.3 Lca062
10.3 Lca485
14.3 Lca434
36.6 LcaTe0409
52.6 Lca440
LG7
0.0 LcaTe0395
5.4 LcaTe0484
17.0 Lca377
22.2 Lca918
LG8
0.0 LcaTe0264
5.2 Lca266
7.5 Lca086
14.2 Lca993
21.9 LcaTe0023
30.1 Lca606
LG9
0.0 Lca249
22.3 Lca301
36.5 LcaE201
42.4 Lca454
44.8 LcaB155
LG10
0.0 Lca212
6.9 LcaB201
8.5 Lca590
13.8 Lca472
16.9 LcaTe0070
19.3 Lca002
21.2 LcaTe0242
23.0 Lca351
31.6 Lca525
39.8 Lca405
LG11
0.0 Lca063
34.1 LcaTe0027
47.0 Lca058
48.8 LcaE232
50.8 LcaE233
55.5 Lca518
LG12
0.0 LcaTe0455
18.2 LcaTe0499
29.5 Lca907
65.1 Lca178
82.8 Lca188
LG13
0.0 Lca598
1.7 Lca146
3.3 LcaTe0232
7.0 LcaE113
27.0 Lca615
27.9 Lca1027
32.5 LcaTe0418
LG14
0.0 Lca147
5.9 LcaTe0465
7.3 LcaTe0011
11.1 LcaB055
LG15
0.0 Lca534
9.9 Lca621
11.1 Lca256
13.3 Lca069
20.4 Lca163
31.8 Lca271
LG16
0.0 LcaTe0223
19.9 Lca020
38.4 Lca283
LG17
0.0 LcaTe0064
1.7 Lca021
9.4 LcaB051
13.7 Lca1012
15.8 Lca057
17.6 LcaB023
20.8 Lca628
LG18
0.0 Lca530
11.0 Lca519
22.4 LcaTe0204
46.6 Lca938

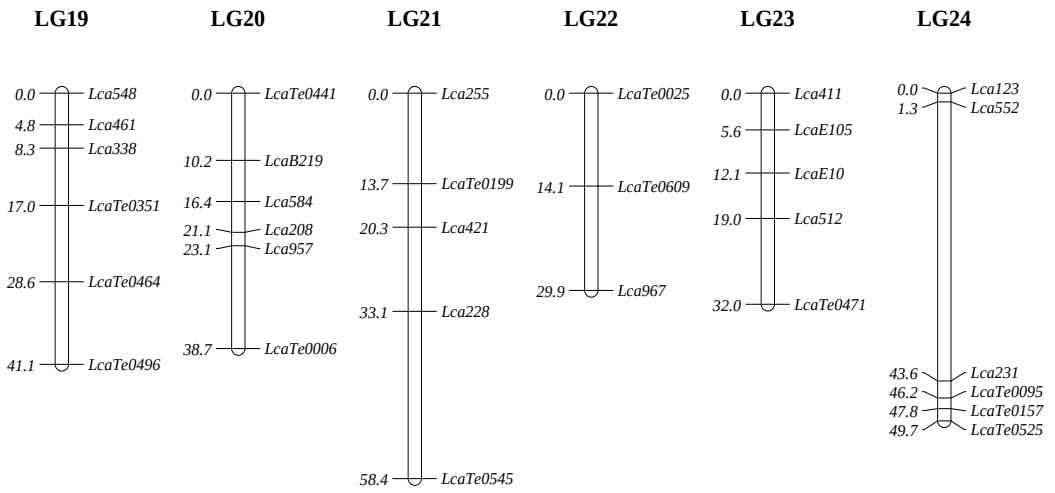


Fig. 2.4 The re-constructed linkage map consisting of 24 linkage groups of Asian seabass. Each vertical bar represents a linkage group. The numbers on the left of each bar represent the locations in cM of the markers in the linkage group. The texts on the right of the bar represent the names of said markers.

2.3.4. QTL mapping

One of the main functions of construction of a linkage map is to identify chromosome regions and/or genes that are responsible for the trait of interest by determining the association between phenotype and genotype of markers, which is referred to as QTL mapping (Collard *et al.* 2005; Gjedrem & Baranski 2010; Boopathi 2012). In this study, the re-constructed linkage map was used to link the genotypes of markers to the traits of VNN resistance and survival time using MapQTL 6 (Van Ooijen & Kyazma 2009). QTL mapping detected 13 (Table 2.2) and 10 (Table 2.3) QTL for VNN resistance and survival time, respectively. Four significant QTL at genome-wide level located in LGs 20, 21 and 24, were detected for resistance. PVE, in percentages, attributed to those QTL were 4.10, 3.60, 3.50 and 3.70%, respectively. Moreover, one significant QTL at chromosome-wide level, located in LG 10, was also detected for resistance with a PVE of 3.5%, which was slightly lower than the PVE of three significant QTL at genome-wide level. In addition, eight suggestive QTL were identified in seven LGs with PVE ranging from 2.2 to 2.9% (Table 2.2). For the survival time, five significant QTL (one in LG 9, two in LG

20 and two in LG 21) were identified at genome-wide level (Table 2.3). They explained 3.00-3.30% of the phenotypic variance. Additionally, five suggestive QTL (located in LGs 4, 10 and 24) were identified at chromosome-wide level with PVE varying from 2.20-2.90%. There was no QTL detected for survival time at genome-wide level.

Table 2.2 QTL detected for resistance to VNN in Asian seabass fingerlings.

LG	Interval (cM)	Sig.	Threshold LOD	Peak LOD	Peak position (cM)	PVE (%)	Nearest marker	Marker position	K*	Sig.
2	0.00-6.73	suc	c0.05 = 2.6	3.02	6.73	2.60	LcaTe0360	6.73	6.63	**
2	11.59-11.59	suc	c0.05 = 2.6	2.65	11.59	2.30	LcaTe0359	13.58	10.18	**
4	26.45-32.30	suc	c0.05 = 2.4	2.76	29.30	2.40	LcaTe0075	29.30	12.08	****
6	34.34-36.60	suc	c0.05 = 2.5	2.56	36.60	2.20	LcaTe0409	36.60	11.86	***
10	8.52-8.52	sic	c0.01 = 3.4	3.98	8.52	3.50	Lca590	8.52	18.32	******
14	0.00-4.00	suc	c0.05 = 2.1	2.45	0.00	2.20	Lca147	0.00	10.63	**
15	20.33-29.37	suc	c0.05 = 2.4	2.66	24.37	2.30	Lca163	20.37	13.30	****
20	0.00-3.00	sig	g0.05 = 4.0	4.69	0.00	4.10	LcaTe0441	0.00	19.52	*******
20	27.13-34.13	suc	c0.05 = 2.4	2.88	30.13	2.50	Lca957	23.13	6.22	**
21	18.71-31.32	suc	c0.05 = 2.6	3.33	24.32	2.90	Lca421	20.32	15.58	****
21	58.12-58.37	sig	g0.05 = 4.0	4.16	58.37	3.60	LcaTe0545	58.37	18.21	******
24	33.30-43.30	sig	g0.05 = 4.0	4.08	38.30	3.50	Lca231	43.61	4.99	**
24	44.61-47.77	sig	g0.05 = 4.0	4.20	47.20	3.70	LcaTe0157	47.77	6.06	**

Note: Sig.: significant level. suc: suggestive; sig: significant. LOD: log of odds. PVE: phenotypic variation explained. K: Kruskal-Wallis test. QTL are suggestive at chromosome-wide (suc) or significant at chromosome-wide (sic) or significant at genome-wide (sig). ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.005$, ***** $p < 0.001$, ****** $p < 0.0005$, ******* $p < 0.0001$.

Table 2.3 QTL detected for survival time to VNN in Asian seabass fingerlings.

LG	Interval (cM)	Sig.	Threshold LOD	Peak LOD	Peak position (cM)	PVE (%)	Nearest marker	Marker position	K*	Sig.
4	24.45-35.30	suc	c0.05 = 2.4	2.82	29.30	2.50	LcaTe0075	29.30	13.67	****
9	21.00-24.31	sic	c0.01 = 3.1	3.44	22.31	3.00	Lca301	22.31	10.17	**
10	0.00-6.86	suc	c0.05 = 2.6	3.15	4.00	2.70	LcaB201	6.86	17.31	*****
10	8.52-9.52	suc	c0.05 = 2.6	3.33	8.52	2.90	Lca590	8.52	17.31	*****
10	14.76-17.87	suc	c0.05 = 2.6	2.81	16.76	2.50	LcaTe0070	16.87	13.05	****
20	0.00-3.00	sic	c0.01 = 3.2	3.77	0.00	3.30	LcaTe0441	0.00	15.62	*******
20	28.13-32.13	sic	c0.01 = 3.2	3.44	30.13	3.00	Lca957	23.13	9.19	**
21	20.32-30.32	sic	c0.01 = 3.3	3.71	25.32	3.20	Lca421	20.32	17.81	******
21	38.12-50.12	sic	c0.01 = 3.3	3.61	45.12	3.1	Lca228	33.12	12.33	****
24	46.20-47.77	suc	c0.05 = 2.4	2.49	47.20	2.20	LcaTe0157	47.77	4.88	**

Note: Sig.: significant level. suc: suggestive. sig: significant. LOD: log of odds. PVE: phenotypic variation explained. K: Kruskal-Wallis test. QTL are suggestive at chromosome-wide (suc) or significant at chromosome-wide (sic). ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.005$, ***** $p < 0.001$, ****** $p < 0.0005$, ******* $p < 0.0001$.

These results suggest that genetic architecture for VNN disease resistance in the examined family is polygenic and controlled by many loci with small effects. Similarly, 10 QTL with up to 27% PVE were found for resistance against the parasite *Gyrodactylu salaries* in Atlantic salmon (Gilbey *et al.* 2006). Additionally, resistance against the parasite *P. marinus* in eastern oyster was controlled by twelve QTL (Yu & Guo 2006). It seems that vast majority of evidence appears to favor the notion that disease resistance is a complex trait. However, contrary to these findings, several pieces of evidence challenged that disease resistance could also be a result of major QTL with large effect. For example, two independent groups reported that resistance against IPN in Atlantic salmon was affected by two major QTL with PVE of more than 23% (Houston *et al.* 2008; Moen *et al.* 2009). One of the two groups further demonstrated 50.9% of PVE of susceptibility to freshwater IPN could be explained by a single QTL in Atlantic salmon (Houston *et al.* 2010). Other evidence also pointed to a major QTL contributing 50% of PVE for resistance against the viral LD in Japanese flounder (Fuji *et al.* 2006). Besides viral diseases, resistance against bacteria was also mapped to a major QTL with a PVE of 50-86% across families in rainbow trout (Baerwald *et al.* 2011). Furthermore, a major QTL with a PVE of 32.9-35.5% for resistance against parasitic disease was recently identified in yellowtail (Ozaki *et al.* 2013). Altogether, these data suggest that disease resistance is a complicated genetic issue. The locations and effects of QTL for disease resistance depend on species, types of pathogen and environmental factors.

Further analysis showed that one significant QTL was detected for both VNN resistance and survival time, five other QTL were detected at approximately the same location in different LGs and considered significant or suggestive for both VNN resistance and survival time (Table 2.2 and 2.3). For example, two QTL located in LGs 10 and 24 were identified as significant for resistance and suggestive for survival time; two QTL

located in LGs 20 and 21 were suggestive for resistance and significant for survival time. One consistent QTL in LG 4 was identified as suggestive for both VNN resistance and survival time. Similarly, in other fish species, several QTL for VNN resistance and survival time in close positions within same LGs were also reported in turbot against furunculosis caused by bacteria pathogen *A. salmonicida* (Rodriguez-Ramilo *et al.* 2011), scuticociliatosis caused by a parasite *P. dicentrarchi* (Rodriguez-Ramilo *et al.* 2013) and VHS caused by novirhabdovirus (Rodriguez-Ramilo *et al.* 2014).The possible reasons for concordance include: 1) the high correlation between both traits as the resistance includes survival time, and 2) the same genes being involved in the underlying mechanism for disease resistance. It is also possible that the two measurements may actually reflect different aspects of disease resistance (Ødegård *et al.* 2011a). However, my data suggest that it is impossible to differentiate between VNN resistance and survive time. Further fine mapping of QTL, followed by identification and characterization of genes located in those QTL, may resolve the issue.

Besides the consistent QTL detected for both VNN resistance and survival time, seven significant and suggestive QTL in seven LGs were found only for resistance and four QTL in three LGs were only for survival time. Similar results were also reported in QTL for resistance against *A. salmonicida* (Rodriguez-Ramilo *et al.* 2011), *P. dicentrarachi* (Rodriguez-Ramilo *et al.* 2013) and VHS (Rodriguez-Ramilo *et al.* 2014) in turbot. These results could jointly imply that there could be some specific genes responsible for VNN resistance and survival time. Additionally, it should be noted that several LGs contain more than one QTL for both VNN resistance and survival time. The possible reason could be that those disease resistance genes underpinning the QTL might be enriched in several chromosomes and involved in the same pathway against viral infection (Baerwald *et al.*

2011). However, without the sequence information of those chromosomes, this remains as a speculation.

One significant QTL with the highest PVE, spanning 3 cM in LG 20 (Table 2.2 and 2.3, Fig. 2.5), was detected for both VNN resistance and survival time, suggesting that this QTL might have an influence on both VNN resistance and survival time. The marker LcaTe0441 located in this QTL may be useful in the selection of fish with resistance to VNN. Certainly, before utilization of the marker in MAS, it is essential to examine the detected QTL in other families and populations. Furthermore, dissection of this region may lead to isolation of the corresponding genes for disease resistance. Based on estimation of a reference linkage map (Wang *et al.* 2011), the genetic distance approximately equals 0.97 Mb in the physical map, and could contain dozens or even hundreds of genes. Therefore, future studies should focus on fine mapping this QTL by genotyping more markers in larger populations, which could increase the power to detect more recombination in this region.

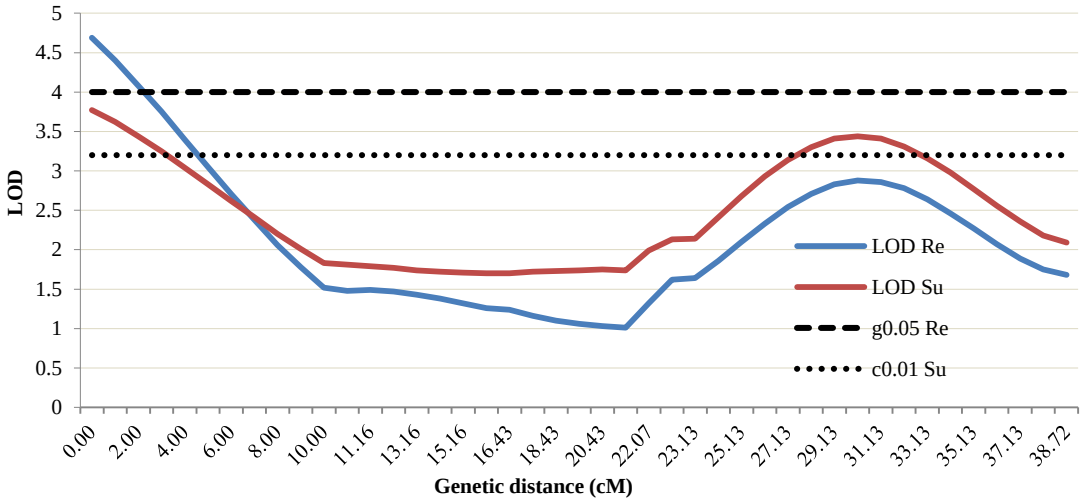


Fig. 2.5 QTL for VNN resistance and survival time in LG 20 of Asian seabass fingerlings. Threshold of LOD value, g0.05 for genome-wide significance at 0.05 level, c0.01 for chromosome-wide significance at 0.01 level for resistance (Re) and survival time (Su).

For this study, some limitations in this experiment were noticed. Firstly, phenotypic divergence of parents might be a concern affecting the detection power of QTL

against VNN in Asian seabass. The survivors from either artificial challenging experiment or natural outbreaks of disease especially viral diseases are commonly excluded as stock brooders with fears that they might be disease carriers having the potential to transmit diseases (Vike *et al.* 2009). Absence of trait evaluation about founder parents could pose a challenge to Asian seabass by compromising the detection power of QTL against VNN. Secondly, using single family may have higher power to detect family specific or rare QTL, it could also trade off the detection power of common QTL (Li *et al.* 2011). In contrast, approach of multi-family mapping permits cross-evaluation and validation of QTL in multiple genetic backgrounds. Thus, could enhance the detection power for shared QTL, improve the resolution preciseness of QTL positions and estimation accuracy of QTL effects (Sneller *et al.* 2009). Consequently, efforts will be made to challenge multi-family of Asian seabass with NNV and verify the detected QTL in these families, as well as mapping common QTL against VNN in multi-family. Lastly, the density in this map is proper for the primary QTL mapping as a single family mapping population allows a relatively large detection power. However, these mapped QTL have a relative large confidence interval, which could hinder the identification of corresponding genes underlying these QTL. Therefore, further research could focus on the sequencing-based marker discovering and genotyping technology such as GBS, which will be presented in Chapter 3.

2.4. Conclusion

In the present study, the linkage map of Asian seabass consisting of 145 microsatellites was re-constructed. Using this map, five significant and eight suggestive QTL for resistance, five significant and five suggestive QTL for survival time, were mapped for VNN resistance and survival time to VNN in Asian seabass, respectively. Multiple QTL identified in this study indicate that VNN resistance is polygenetic. This

could help us understand the genetic architecture of disease resistance, and provide valuable information for fine mapping.

Chapter 3. Fine mapping QTL for resistance to VNN disease using a high-density map in Asian seabass

3.1. Introduction

A large number of polymorphic genetic markers are essential for linkage map construction and QTL mapping. Currently, most linkage maps in aquatic species are built on genetic markers, such as RFLPs, AFLPs and SSRs, and a few of them are constructed on SNPs or mixed types of markers (Yue 2014). To date, linkage maps of over 45 fish species have been reported (Yue 2014). However, development of these markers based on Sanger sequencing and genotyping large populations are slow, labor-intensive, time-consuming, and expensive. Moreover, the resolution of linkage maps built on these markers is moderate. Precise QTL mapping and further determination of causative polymorphisms, even positional cloning of the corresponding genes, require a great number of genome-wide polymorphic markers and high-resolution linkage maps to saturate the LD between marker and QTL (Goddard & Hayes 2009). It is very hard for such conventional methods for marker development and genotyping to meet the increasing requirement for robust development of a large number of unbiased markers across the whole genome and high-density genotyping a large population at low cost (Davey *et al.* 2011).

With the rapid advances of NGS technology, high-throughput sequencing-based marker discovery and genotyping have been quickly developed and adopted in many organisms (Davey *et al.* 2011; Poland & Rife 2012). The sequencing-based genotyping methods generally include multiplexing a proper number of samples with barcodes to greatly reduce per sample cost, and only sequencing a small fraction of a genome with increasing times of coverage, which naturally improves the accuracy of genotyping (Davey

et al. 2011). A small portion of a genome can be archived by reducing the complexity of the genome by restriction enzymes, such as the RRLs, RAD-seq and GBS (Davey *et al.* 2011). Among them, GBS is a widely used approach especially in the non-model organisms without genome references or solid genomes (Poland & Rife 2012). The improved GBS approach employs two enzymes, one common cutter and one rare cutter. The combination of two enzyme cutters enables this approach to capture fragments associated with the rare cutters, which are roughly evenly distributed across the genome (Poland & Rife 2012). The hundreds of thousands fragments are then sequenced and millions of reads are produced, generating tens of thousands of unbiased SNPs spaced across the whole genomes. The feature of producing a larger amount of unbiased markers in an inexpensive way, enables GBS to become the preferable approach to build high-density and high-resolution maps, facilitating QTL mapping and genomic selection, even map-based cloning (Poland & Rife 2012). Riding the wave of GBS, high-density maps and/or QTL mapping for economically important traits have been conducted in several aquatic species, including Atlantic salmon for disease resistance (Houston *et al.* 2012) and linkage map (Gonen *et al.* 2014), blue catfish (*Ictalurus furcatus*) for genetic structure (Li *et al.* 2014), sea cucumber for body weight (Tian *et al.* 2015), and Pacific white shrimp (Yu *et al.* 2015), Asian seabass (Wang *et al.* 2015a) and large yellow croaker (Xiao *et al.* 2015) for growth traits. These works confidently demonstrate that GBS has the ability to construct high-density maps, which facilitate QTL mapping.

To fine map these QTL and further identify candidate genes underlying the QTL, GBS was employed to generated6425 SNPs. A high-density linkage map consisting of 2852 SNPs, and 148 SSR markers from a previous study was constructed. QTL mapping analysis was further conducted and each four moderate QTL for VNN resistance and survival time, with PVE ranging from 7.8 to 11.0%, were identified. Protocadherin alpha-

C 2-like (*pcdhαc2*), was further identified as a candidate gene underpinning a QTL with largest PVE. Further association study in multiple families showed that a six bp indel in the second intron of *pcdhαc2* was significantly associated with the VNN resistance. Furthermore, its expression levels in the brain, kidney, muscle and skin were significantly up-regulated in the NNV challenged Asian seabass. The results could facilitate MAS in selective breeding schemes for disease resistance and lay the foundation for further detailed functional analysis of the potential candidate genes for VNN resistance in Asian seabass.

3.2. Materials and methods

3.2.1. Mapping population

The mapping population used in this study was originally derived from a population challenged with NNV as described in Chapter 2. The first 95 dead and 95 survived fish plus two parents were selected to form a panel for downstream analysis.

3.2.2. DNA isolation using high salt precipitation method

The above-mentioned panel was used to extract high quality DNA. The quality of DNA extracted using the approach developed by our lab (Yue & Orban 2005) was relatively low, with some degradation and protein contamination. These DNA did not meet the requirement for construction of sequencing libraries as the latter required relatively high quality DNA (clean and no clear degradation). Therefore, a new approach using the salt precipitation method (Aljanabi & Martinez 1997) was adopted to extract genomic DNA of the panel. In brief, about 0.25 mm^2 of fin tissue was cut and ethanol was removed. The fin cutting was homogenized in 800 µl salt homogenizing buffer containing proteinase K (final concentration of 400 µg/ml) at 55°C and 160 *x* g for two hours. After that, 600 µl of 6 M NaCl was added to the sample lysate and mixed well through vortexing. The mixture

was centrifuged at 14000 *x* g for 30 min at 4°C. The supernatant was transferred to a new tube before adding an equal volume of ice-cold isopropanol. The mixture was incubated at -20°C for 1 h before centrifugation at 14000 *x* g for 20 min at 4°C. To wash DNA pellet, 1 ml of 70% ethanol was added and the mixture was centrifugation at 14000 *x* g for 10 min at 4°C. Subsequently, ethanol was removed and DNA pellet was dried at room temperature for 20 min before adding 200 µl of sterile water. DNA pellet was incubated at 4°C overnight to ensure complete dissolution of DNA. To assess DNA quality, 2 µl of DNA was electrophoresed on 2% agarose gel stained with 1% EB and visualized using Gel DOC XR (Bio-Rad Laboratories, Hercules, USA).

The concentrations of genomic DNA were determined by plate reader of Infinite® M1000 PRO (Tecan, Männedorf, Switzerland) using Qubit® dsDNA HS Assay Kit (Life Technologies, Carlsbad, USA) following the manufacturer's instructions. In brief, the Qubit® working solution was prepared by diluting dsDNA HS reagent 200 times with dsDNA HS buffer. The three standards were prepared by mixing 200, 195 and 190 µl of working solution with 0, 5 and 10 µl of standard solution, respectively. Additional 198 µl of working solution was mixed with 2 µl of original DNA in a 96-well plate. The assay plate was run on plate reader of Infinite® M1000 PRO for signal detection. The signal was converted to value of DNA concentration using the three standards. DNA was diluted to a final concentration of 10 ng/µl as a working stock and kept in -80°C for downstream analysis.

3.2.3. Sequencing library preparation and next generation sequencing

The sequencing RAD libraries were prepared using double digestion RAD-seq method with some modifications (Peterson *et al.* 2012) as described in the previous paper (Wang *et al.* 2015a). In brief, 200 ng of each DNA was digested with 20 units of restriction

enzyme of PstI-HF and MspI (New England Biolabs, Ipswich, USA) at 37 °C for 2.5 hours. The DNA fragments were examined by electrophoresis on a 2% agarose gel before ligation with barcoded adaptors (Peterson *et al.* 2012). The ligation products were pooled, followed by the size selection of 350 to 600 bp using Pippin Prep (Sage Science, Beverly, USA), and then a cleanup using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). The total fragments were PCR amplified using Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA), followed by a second clean up using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Quantification of the libraries was determined using KAPA Library Quantification Kits (Kapa Biosystems, Wilmington, MA, USA) by qPCR in the MyiQ Thermal Cycler (Bio-Rad Laboratories, Hercules, USA). The libraries were sequence on NextSeq 500 platform (Illumina, San Diego, USA) to generate raw sequencing single-end reads of 151 bp.

3.2.4. Processing of NGS reads, identifying and genotyping of SNPs

The raw sequencing reads were processed by the program *process_radtags* implemented in Stacks package (version 1.21) (Catchen *et al.* 2011) to remove low quality reads and any uncalled base. To reduce the sequencing errors at the end of each read, all the clean reads were trimmed to 95 bp, following a final step of de-multiplexing bioinformatically and assigning clean reads to each sample. All the downstream analysis of stack assembly, sequence mapping, SNP calling and genotyping were performed by the Stacks platform (Catchen *et al.* 2011) with parameters described in (Wang *et al.* 2015a). For the two parents, the stacks and catalogue loci were constructed with a minimum of 20 times coverage (Catchen *et al.* 2011). For the offspring, a minimum of five times coverage was applied to assemble the stacks. SNP calling and genotyping were conducted by *sstacks*

and *genotypes* (Catchen *et al.* 2011), respectively. Any SNP with more than 20% missing data in both genotype and individual were removed from further analysis.

3.2.5. Construction of a high-density linkage map

All the SNP and SSR markers were submitted to JoinMap 4.1 (Van Ooijen 2011) to construct a linkage map. The process followed the procedure described in Chapter 2.

3.2.6. High-resolution QTL mapping

After linkage analysis, identification and mapping of QTL were carried out by MapQTL 6 (Van Ooijen & Kyazma 2009). The process followed the procedure described in Chapter 2.

3.2.7. Possible candidate gene identification and polymorphism analysis

Following the identification of QTL regions, the nearest marker to the peak of QTL was determined and the sequence harboring the SNP or SSR was retrieved. The obtained sequence was used as seed to retrieve 300 kb sequences from the Asian seabass genome, followed by blasting the Asian seabass transcriptome assembled in Chapter 4 against those sequences by GMAP (Wu & Watanabe 2005) with default parameters to identify genes in these QTL. Genes with the highest score and longest matching length were selected as the candidate genes and their mRNA sequence was retrieved from transcriptome of Asian seabass. UTRs, exons and introns of the candidate gene were determined by Spidey, which is an mRNA-to-genomic alignment program (http://www.ncbi.nlm.nih.gov/spidey/index.html). To identify the polymorphisms in the candidate gene, primer pairs were designed using DNASTAR Lasergene 11 (DNASTAR, Madison, USA) to amplify the whole sequence. PCR products were sequenced using

Sanger sequencing method. In brief, the first round of PCR followed the procedure as described in Chapter 2. To assess PCR products, 5 µl of RNA was electrophoresed on 2% agarose gel stained with 1% EB and visualized using Gel DOC XR (Bio-Rad Laboratories, Hercules, USA). Only the PCR product with a single band was selected for the following a second round of sequencing PCR. The sequencing PCR was performed in a reaction volume of 20 µl, containing 3 µl of 5x sequencing buffer, 2 µl of BigDye® Terminator v3.1 (Thermo Fisher Scientic, Waltham, USA), 2 µM of either forward or reverse primer, 2 µl of first round PCR product and 12.8 µl of water. The thermal cycling conditions were as follows: 26 cycles of 94°C for 30 s, 50°C for 15 s and 60°C for 4 min. The PCR product was cleaned and sequenced by 3730xl DNA analyzer (Applied Biosystems, California, USA). The sequences were analyzed by Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, USA). Low quality nucleotides were removed and high quality nucleotides were used to align to the genomic sequence of candidate gene to identify the SNPs and indel. Microsatellites in the candidate gene were identified using GRAMENE with default parameters (Temnykh *et al.* 2001). Schematic representation of candidate gene was produced by Exon-Intron Graphic Maker (http://wormweb.org/exonintron). All Sanger sequencing PCR and sequence assembly used followed this procedure.

3.2.8. Challenge three-month old Asian seabass with NNV

To extract RNA for determination of expression patterns of the genes of interest, three-month old Asian seabass were transferred from the MAC, AVA in St. John Island, to Temasek Life Sciences Laboratory. Prior to challenging, fish were acclimatized in two tanks with 25 liters of circulated seawater (30°C, salinity 30 ppt, pH 7.6) and saturated oxygen for 5 days. In the whole period of experiment, fish were fed with a commercial feeder (Marubeni Corporation, Tokyo, Japan) twice a day, and half of the seawater was

replaced with fresh seawater every two days. On the day of challenge, three fish were randomly selected and weighed with an average body weight was 16.36 ± 3.04 g. During the challenge, one group of fish was each intraperitoneally injected with 0.1 ml of NNV stock with a concentration of 3.75×10^8 $TCID_{50}$, and designed as the NNV-challenged group; the other, the mock group, was each intraperitoneally injected with an equivalent amount of 0.1 ml of used L-15 medium. All the fish were closely monitored in the whole experimental period. At 5 dpc, three fish from each of the two groups (NNV-challenged and mock) were scarified, before collection of 10 tissues and organs: brain, eye, fin, heart, intestine, kidney, liver, muscle, skin and spleen, respectively.

3.2.9. RNA isolation and reverse transcription

To determine expression levels of candidate genes underlying the identified QTL in Asian seabass by quantitative reverse transcription PCR (qRT-PCR), total RNA was isolated from these tissues and organs using TRIzol (Life Technologies, Carlsbad, USA) following the manufacturer's instructions. The whole process of RNA isolation, digestion with DNase I and reverse-transcription all followed the procedure described in Chapter 2. The cDNA samples were diluted 5x with Nuclease-free water and kept at -80°C for downstream analysis.

3.2.10. Determination of expression pattern of candidate gene using qRT-PCR

To determine the expression pattern of the candidate gene after NNV infection, qRT-PCR was conducted on the cDNA converted from RNA extracted from 10 organs and tissues at 5 dpc in the mock and NNV challenged groups. The ORF of the candidate gene was used to design primer pair. In this study, *pcdhαc2* was the possible candidate gene for QTL qVNN-Re_20.1 and qVNN-Su_20.1 and primer pairs of Lca-pcdhαc2-q (Appendix

Table A2) was obtained. qPCR reactions were performed on Applied Biosystems 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, USA), using SYBR Green as fluorescent dye. A 10 µl qPCR reaction contained 1.6 water, 3 µl (15ng) of 10x diluted cDNA, 0.2 µl (2 µM) of each primer and 5 µl of 2x master mix from KAPA SYBR® FAST qPCR Kits (Life Technologies, Carlsbad, USA). The qPCR followed the conditions of 95°C for 3 min, 40 cycles of 95°C for 10 s and 58°C for 30 s. Each reaction had three repeats. Raw data were converted to cycle threshold (*C*t) values using the software provided by Applied Biosystems (Thermo Fisher Scientific, Waltham, USA). $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen 2001) was used to analyze the relative quantifications and fold change of each gene using the elongation factor 1-alpha 1 (*EF1α1*) as a reference gene (Wang *et al.* 2014d). In brief, $\Delta\Delta Ct = (C_{T,T}, - C_{T,\ EF1a1})_{\text{Treatment V}} - (C_{T,T}, - C_{T,\ EF1a1})_{\text{Treatment M}}$, where Treatment V represented each NNV-challenged sample, Treatment M represented each the corresponding mock sample, $C_{T,T}$ represented *C*t value of the each target gene of either NNV-challenged or mock sample, $C_{T,\ EF1a1}$ represented *C*t value of the reference gene of *EF1a1* of either NNV-challenged or mock sample. The relative expression level of a gene was calculated by $2^{-\Delta\Delta Ct}$. The fold change of each up-regulated gene was the ratio of relative expression of the VNN-challenged to that of mock sample. The fold change of each down-regulated gene was the ratio of relative expression of the mock sample to that of VNN-challenged sample. All qPCR and calculation of gene expression described in this thesis followed this procedure.

3.2.11. Association of the candidate gene with VNN disease in multiple families of Asian seabass

To conduct association of *pcdhαc2* with VNN disease resistance, a primer pair of Lca-pcadhαc2-del was designed to target the region containing the indel (Appendix Table

A2). Forward primer was labeled with fluorescence FAM (Sigma-Aldrich, Missouri, USA) at its 5' end, and genotyping was performed in the association population by PCR and fragment length was determined as described in Chapter 2. The association population was challenged with NNV, following the same procedures as described in Chapter 2. The association population consisted of more than 1127 individuals (651 mortalities and 476 surviving fish) from 43 families produced by a mass cross of 15 parents.

3.2.12. Data deposition

All the raw reads were submitted to the sequence reads archive (SRA), NCBI database with an accession number of PRJNA317915.

3.3. Results and discussion

3.3.1. Discovering of genome-wide SNPs and genotyping

With high efficiency and low cost, NGS technology has revolutionized the way how the polymorphic markers are developed and genotyped (Poland & Rife 2012). Using GBS, four DNA libraries, each library consisting of 45 fish (either early dead or randomly selected surviving fish) and one parent, were sequenced. Therefore, 95 early dead fish, 95 early dead fish and two parents were used. These fish with extreme phenotypes could provide strong QTL mapping power (Georges 2007). After reads processing, including removing low quality reads, trimming and filtering missing genotype and offspring, a total of 784.68 million high-quality clean reads were obtained. Of these filtered reads, an average of 4.36 million reads were assigned to each offspring, 14.74 million and 19.05 million reads were assigned to sire and dam, respectively. Using the clean reads from parents, a catalogue consisting of 18857 loci was obtained. This catalogue was used as a reference to obtain the SNPs and genotypes of the mapping population. A total of 6425

SNPs was identified in 85 dead and 94 surviving fish with the filtering criteria of < 20% missing data across all the samples and > 5 x coverage for each data point. The number of SNPs produced in this study is much higher than the number of 3928 SNPs obtained by a previously studied of Asian seabass (Wang *et al.* 2015a) and 4275 SNPs for blue catfish (Li *et al.* 2014). Similarly, 6146 and 6712 SNPs were reported in Pacific white shrimp for growth traits (Yu *et al.* 2015) and Atlantic salmon for IPN resistance (Houston *et al.* 2012), respectively. In contrast, 7839 and 8257 SNPs were discovered in sea cucumber for growth-related QTL (Tian *et al.* 2015) and Atlantic salmon for conduction of high-density map (Gonen *et al.* 2014). It is worth to note that there is no universal standard for the number of SNPs in QTL related works. The number of SNPs generated by GBS depends on several factors including species, genome size, genome composition, genome information, varieties of GBS platform, sequence depth and stringency of criteria for filtering low quality reads. The sample size of 179 fish for fine mapping QTL in this work is higher than a previous study used 144 individuals for QTL mapping. In that study, it successfully fine mapped a gene of peroxisomal acyl-coenzyme A oxidase 1 (*acox1*) responsible for a major QTL for growth in Asian seabass (Wang *et al.* 2015a), which could indicate the powerful detection ability of GBS.

3.3.2. Construction of a high-density linkage map

A linkage map is essential for downstream analysis like QTL mapping. To construct a linkage map, all the 6425 SNPs discovered from GBS and 154 SSRs (149 from the study in Chapter 2 and 5 newly developed markers based on transcriptome in Chapter 4 were assessed for Mendelian segregation by Joinmap 4.1 (Van Ooijen 2011) before map construction. After removal of distorted markers, 3017 SNPs and 154 SSRs were used to construct a linkage map. Among these, a total of 2852 SNPs and 148 SSRs, were

successfully mapped to the linkage map. This high-density map consisted of 24 LGs contained 3000 high-confidence markers and spanned 2957.79 cM with an average marker interval of 1.28 cM (Tables 3.1 and Fig. 3.1). In this linkage map, a total of 670 markers were observed to be clustered together at 366 positions across the 24 LGs and co-segregated in groups (Table 3.1). No recombination happened between these markers, thus the marker interval was zero. Although these markers were known to be clustered together in the map positions, their orientations were hard to determine. This phenomenon is referred as bin signature (Fig. 3.2) (Milbourne *et al.* 2000). Similar findings of bin signature has also been reported in the linkage map of Pacific white shrimp (Yu *et al.* 2015). Several reasons could contribute to this, including the very close physical positions of these markers in the same chromosome resulting in nearly no recombination. Alternatively, these markers could be physically distant from each other but located in the cold spots of recombination (Van Os *et al.* 2006). Additionally, a relatively small mapping population (e.g. 179 individuals in this study) could limit the detection of recombination events during meiosis. Thus, increasing the number of individuals could increase the power to detect recombination between markers as well as to determine their orientations in the bin signature. Nevertheless, the quality of this map is much higher than the previous map for QTL mapping for NNV resistance in terms of marker number (3000 vs 145) and density (1.28 vs 6.90 cM). This demonstrates that GBS has the ability to robustly generate a larger number of high-quality and high-confidence markers than conventional SSR marker discovery, for construction of linkage maps in aquatic species. This further indicates that the present map has much more power for QTL mapping than the previous SSR marker based map, increasing the capability to capture QTL while reducing the possibility of false positive QTL. The higher resolution of the map could also narrow down the QTL confidence interval. Nevertheless, the number of mapped SNPs was slightly less than the 3321 SNPs produced by GBS in a map for QTL mapping for growth in Asian seabass

(Wang *et al.* 2015a). A possible reason could be due to the differences of cross design. For example, a backcross population of 179 individuals was used while the previous map used an F_2 population of 144 for the map (Wang *et al.* 2015a). It is obvious that backcross could reduce the genetic variance because one fourth of homozygous genotypes were absent in the offspring population compared to the F_2 population. Therefore, this absence has translated into the reduced number of polymorphic SNPs.

In addition, the total length of the present linkage map (2957.79 cM) was much longer and its average marker interval was larger, than those in the previous map of 1577.67 cM and 0.52 cM (Wang *et al.* 2015a), respectively. The linkage map is built on chromosome recombination during meiosis. There is a common difference in recombination frequencies between sexes of fish species, including Asian seabass, in which the length of female LGs was longer than that of male (Wang *et al.* 2011; Wang *et al.* 2015a). The longer the linkage map is, the less recombination, under a comparable number of genetic markers in the same species. However, the reasons for the huge difference in total length between the two linkage maps in Asian seabass remain unclear. It could be due to the different families used for constructing the linkage maps and further linkage map analysis on more families of Asian seabass could clear this suspicion.

The length of each LG ranged from 80.70 (LG15) to 180.30 (LG8) cM with an average length of 123.24 cM (Table 3.1). The number of markers in each LG varied from 43 (LG6) to 233 (LG21) with an average number of 125. The marker interval of each LG ranged from 0.54 (LG19) to 7.63 (LG18) cM with an average of 1.28 cM. This small average marker interval could substantially improve the map resolution, which naturally enhances the effectiveness of fine mapping. The largest marker interval gap was 38.95 cM, located in LG 20. Furthermore, marker intervals were not consistent across all the LGs. Comparing this map with the previous one (Wang *et al.* 2015a) which was constructed by

markers generated by the same method of GBS, found that there were differences in length, number of markers and average marker interval in each LG. This could be a result of different families, different genomic structure of the chromosomes and/or the specific sequences produced by enzyme digestion. Further studies of mapping those reads to the genome reference of Asian seabass could answer this question.

Table 3.1 Statistics of 24 linkage groups in the linkage map of Asian seabass.

LG	N. of markers	N. of markers in bin signature	N. of markers for map statistics	Total length (cM)	Mark interval (cM)
1	130	23	107	163.19	1.54
2	124	5	119	142.59	1.21
3	233	79	154	90.79	0.59
4	131	11	120	150.39	1.26
5	230	66	164	88.12	0.54
6	159	28	131	89.20	0.69
7	173	36	137	93.02	0.68
8	224	82	142	180.30	1.28
9	130	23	107	127.60	1.20
10	177	32	145	125.17	0.87
11	63	11	52	116.67	2.29
12	122	22	100	88.45	0.89
13	93	18	75	143.93	1.95
14	59	40	19	137.38	7.63
15	117	29	88	80.70	0.93
16	72	15	57	97.35	1.74
17	92	16	76	118.76	1.58
18	43	6	37	149.24	4.15
19	214	58	156	83.17	0.54
20	66	14	52	96.65	1.90
21	78	18	60	148.11	2.51
22	65	15	50	166.96	3.41
23	112	33	79	132.43	1.70
24	93	19	74	147.64	2.02
total	3000	670	2330	2957.79	-
minimum	43	5	37	80.70	0.54
maximum	233	82	164	180.30	7.63
average	125	28	97	123.24	1.28

Note: LG: linkage group; N. : number; -: not available

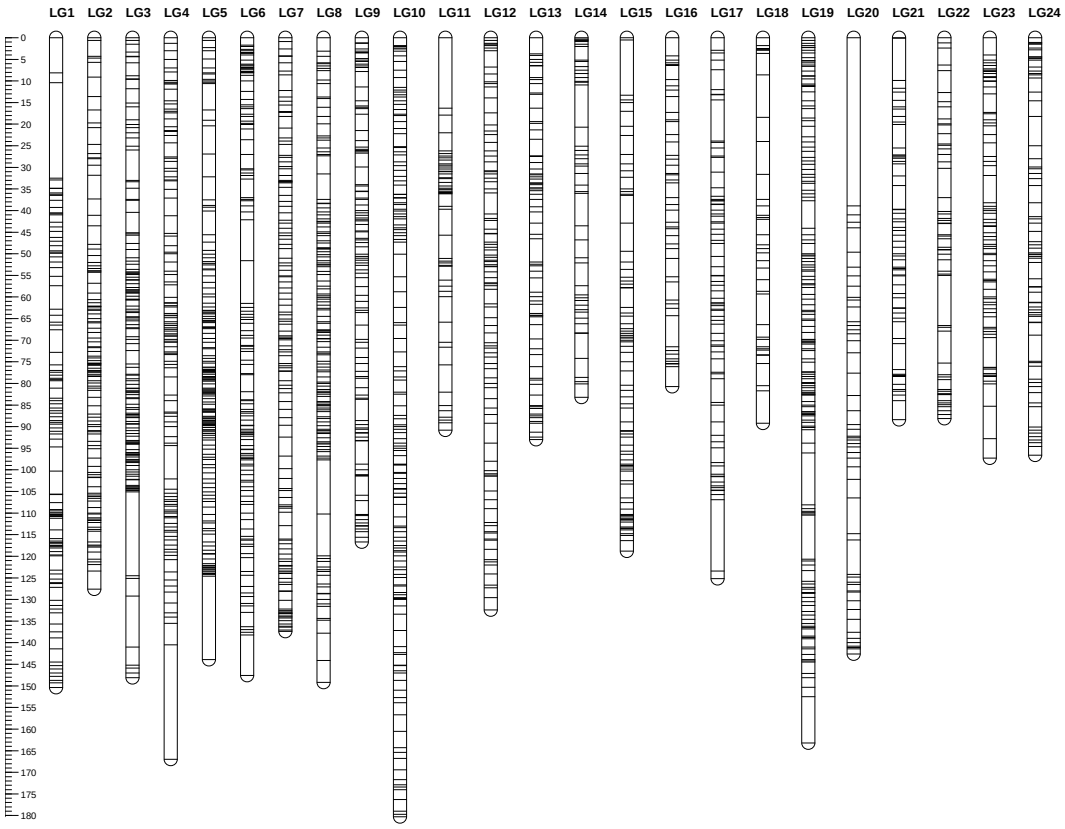


Fig. 3.1 The linkage map of Asian seabass consisting of 24 linkage groups. Ruler measures the length of each linkage group in cM. Each vertical bar represents a linkage group. Each line in the bar represents a marker.

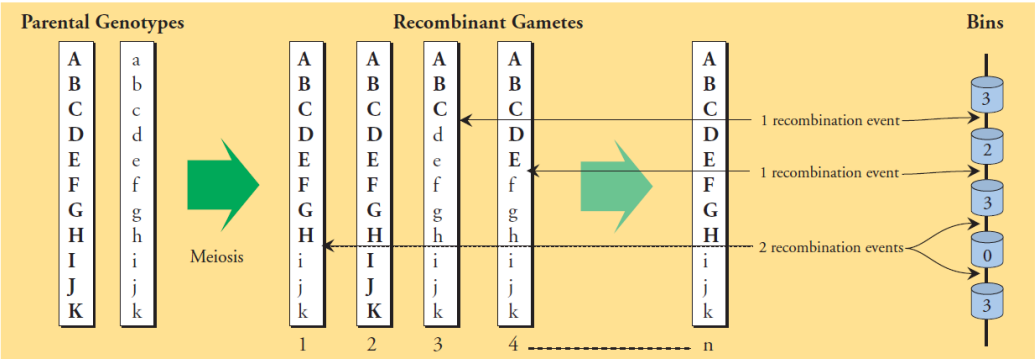


Fig. 3.2 The bin map concept. A complete heterozygous, diploid parental chromosome undergoes meiosis to produce recombinant haploid gametes. Following the segregation of marker alleles (capital and small letters) in progeny derived from these gametes allowed the generation of a linkage map. Markers not separated by a recombinant event (population size of n) co-segregated, and were placed in the same co-segregation bin. When more than one recombination events occurred between two consecutive markers, empty bin (bin 4 above) were placed in the map to represent every recombination event which could not be visualized. Reprinted from Milbourne and his team (Milbourne *et al.* 2000).

3.3.3. QTL mapping

A high-resolution linkage map could improve the capability to map high-confidence QTL and narrow down the large interval of QTL to a relatively small one, facilitating identification of causative polymorphisms responsible for the QTL. To identify QTL related to NNV resistance in Asian seabass, QTL mapping using the high-resolution map and trait values for VNN resistance and survival time was performed. QTL mapping analysis resulted in four QTL being detected in three LGs (4, 10 and 20) for resistance (Table 3.2), and another four QTL being mapped in four LGs (4, 10, 20 and 23) for survival time (Table 3.3). In contrast, the previous results in Chapter 2 showed that 13 QTL in nine LGs and 10 QTL in six LGs were identified for VNN resistance and survival time, respectively. The differences could be caused by the low-density map, which may have resulted in possible false positives in the previous study (Mackay *et al.* 2009). This also highlights that a high-density map could yield more creditable QTL. For resistance, *qNNV-Re_20.1* located in LG 20 was detected as significant with a PVE of 11.0% (Table 3.2, Fig. 3.3 and 3.7), the highest among all the detected QTL. It was the same QTL detected in the previous study which also explained the highest proportion of phenotypic variance 4.1%. In this study, this QTL spanned 1.76 cM from 76.85 to 78.61 cM with peak position at 77.61 cM, where the SSR marker LcaTe0441 is located, in LG 20 (Table 3.2, Fig. 3.3 and 3.7). In contrast, this QTL spanned a region of 3 cM in the previous map, much larger than in the current study. This clearly demonstrates that the high-density map of the current study has dramatically narrowed down the QTL region to a small confidence interval while increasing the PVE explained by the same QTL. In addition, two more suggestive QTL were detected in LG 10 with PVEs of 9.2 and 8.3% (Table 3.2, Fig. 3.4 and 3.7). The interval of *qNNV-Re_10.1* was 9.8 cM from 60.77 to 70.57 cM with SNP marker 24304 at its peak position of 62.35 cM, while that of the other QTL *qNNV-Re_10.2* was 1.13 cM

from 115.34 to 116.47 cM with SNP marker 25617 at the peak position of 115.63 cM (Table 3.2, Fig. 3.4 and 3.7). One suggestive QTL *qNNV-Re_4.1* with a PVE of 8.5% was detected in LG 4. This QTL region was 1 cM from 41.15 to 42.15 with SSR marker LcaTe0075 at its peak position of 41.15 cM (Table 3.2, Fig. 3.5 and 3.7). This QTL is consistent with the previously identified one with an interval of 5.85 cM and the peak position marker of LcaTe0075. It again shows that the current map could yield more accurate QTL mapping.

For survival time, two significant and other two suggestive QTL were detected in four LGs. One significant QTL *qNNV-Su_20.1*, with the highest PVE of 10.9%, was detected in LG 20 (Table 3.3, Fig. 3.3 and 3.7). It spanned 3.16 cM from 75.85 to 78.61 cM with SSR marker LcaTe0441 at the peak position of 77.61 cM. This QTL was repeatedly identified as significant with the highest PVE for VNN resistance and survival time in both current and previous studies in Chapter 2. It highlights the cross-validation of this QTL by two studies, and strongly suggests that there could be same genes or pathways responsible for both VNN resistance and survival time. This overlapping by VNN resistance and survival time was also noticed by several QTL studies for resistance to viruses (Rodriguez-Ramilo *et al.* 2014), bacteria (Rodriguez-Ramilo *et al.* 2011) and parasites (Rodriguez-Ramilo *et al.* 2013) in turbot (*Scophthalmus maximu*). One significant QTL *qNNV-Su_10.1* with a PVE of 10% was detected in LG10 (Table 3.3, Fig. 3.4 and 3.7). This QTL region spanned 2 cM from 62.35 to 64.35 cM with SNP marker 24304 at the peak position of 62.35 cM. Furthermore, one suggestive QTL *qNNV-Su_4.1* with a PVE of 9.2% was located from 40.10 to 43.15 cM in LG 4, with its peak position at 41.15 cM where the SSR marker LcaTe0075 is located (Table 3.3, Fig. 3.5 and 3.7). One suggestive QTL *qNNV-Su_23.1*, located in LG 23, was determined with a PVE of 7.8%. It spanned 0.33 cM from 78.11 to 78.44 cM with SNP marker 24304 at the peak position of 78.11

(Table 3.3, Fig. 3.6 and 3.7). The different QTL mapped for VNN resistance and survival time could reflect the genes involved different aspects and development stages of disease (Ødegård *et al.* 2011a).

Table 3.2 Identified QTL for resistance (Re) to VNN in Asian seabass.

LG	QTL	Interval (cM)	Sig.	Threshold LOD	Peak LOD	Peak position (cM)	PVE (%)	Nearest marker	Marker position	K*	Sig.
4	qVNN-Re_4.1	41.15-42.15	suc	3.2	3.46	41.15	8.5	LcaTe0075	41.15	13.95	*****
10	qVNN-Re_10.1	60.77-70.57	suc	3.0	3.74	62.35	9.2	24304	62.35	10.878	*****
10	qVNN-Re_10.2	115.34-116.47	suc	3.0	3.38	115.63	8.3	25617	115.63	11.263	*****
20	qVNN-Re_20.1	76.85-78.61	sic	3.9	4.55	77.61	11.0	LcaTe0441	77.61	14.132	******

Note: Sig.: significant level. suc: suggestive. sig: significant. LOD: log of odds. PVE: phenotypic variation explained. K: Kruskal-Wallis test. QTL are significant (sic) and suggestive (suc) at chromosome-wide level. Sig.: significant level ***** p <0.001, ****** p <0.0005

Table 3.3 Identified QTL for survival time (Su) to VNN in Asian seabass.

LG	QTL	Interval (cM)	Sig.	Threshold LOD	Peak LOD	Peak position (cM)	PVE (%)	Nearest marker	Marker position	K*	Sig.
4	*qVNN-Su_4.1*	40.1-43.15	suc	3.1	3.75	41.15	9.2	LcaTe0075	41.15	17.501	******
10	*qVNN-Su_10.1*	62.35-64.35	sic	4.0	4.1	62.35	10.0	24304	62.35	11.44	*****
20	*qVNN-Su_20.1*	75.85-78.61	sic	3.8	4.5	77.61	10.9	LcaTe0441	77.61	10.947	*****
23	*qVNN-Su_23.1*	78.11-78.44	suc	3.0	3.17	78.11	7.8	96495	78.11	0.013	-

Note: Sig.: significant level. suc: suggestive. sig: significant. LOD: log of odds. PVE: phenotypic variation explained. K: Kruskal-Wallis test. QTL are significant (sic) and suggestive (suc) at chromosome-wide level. Sig. significant level ***** p <0.001, ****** p <0.0005

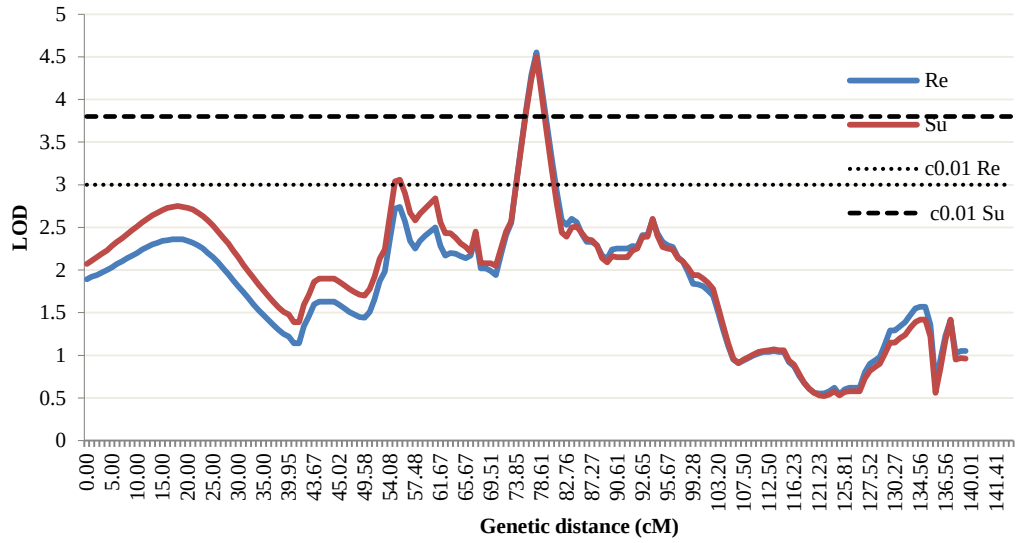


Fig. 3.3 QTL of *qVNN-Re_20.1* and *qVNN-Su_20.1* detected in LG 20 of Asian seabass. Re for resistance and Su for survival time, c0.01 Re and c0.01 Su for 0.01 significant levels on chromosome-wide.

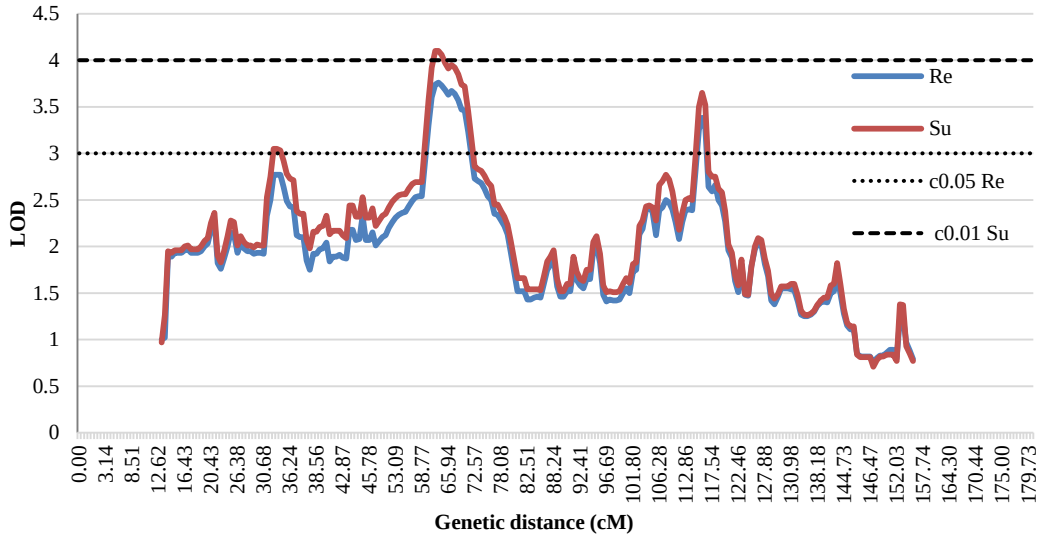


Fig. 3.4 QTL of *qVNN-Re_10.1*, *qVNN-Re_10.2* and *qVNN-Su_10.1* detected in LG 10 of Asian seabass. Re for resistance and Su for survival time, c0.05 Re and c0.01 Su for 0.05 and 0.01 significant levels on chromosome-wide, respectively.

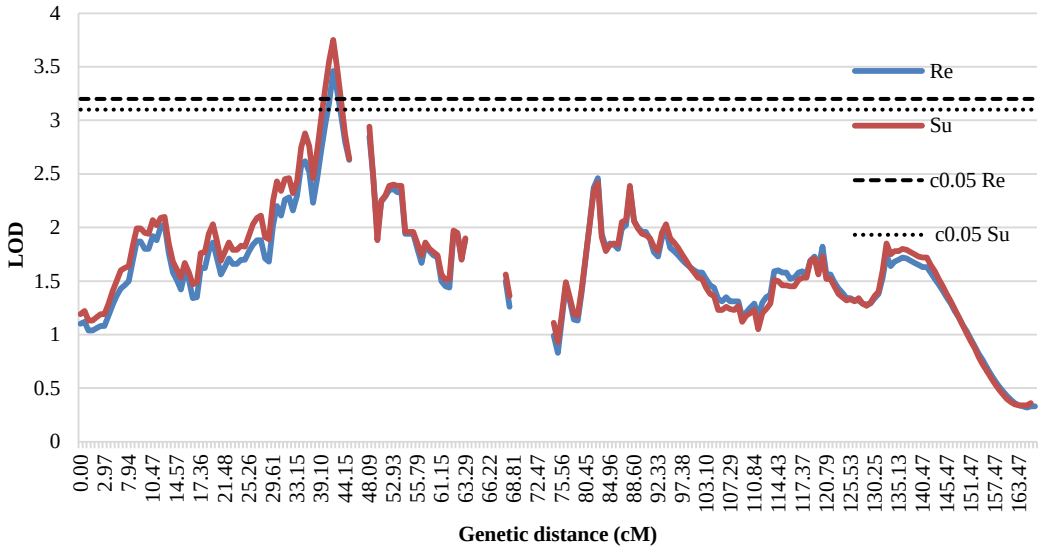


Fig. 3.5 QTL of *qVNN-Re_4.1* and *qVNN-Su_4.1* detected in LG 4 of Asian seabass. Re for resistance and Su for survival time, c0.05 Re and c0.05 Su for 0.05 significant levels on chromosome-wide.

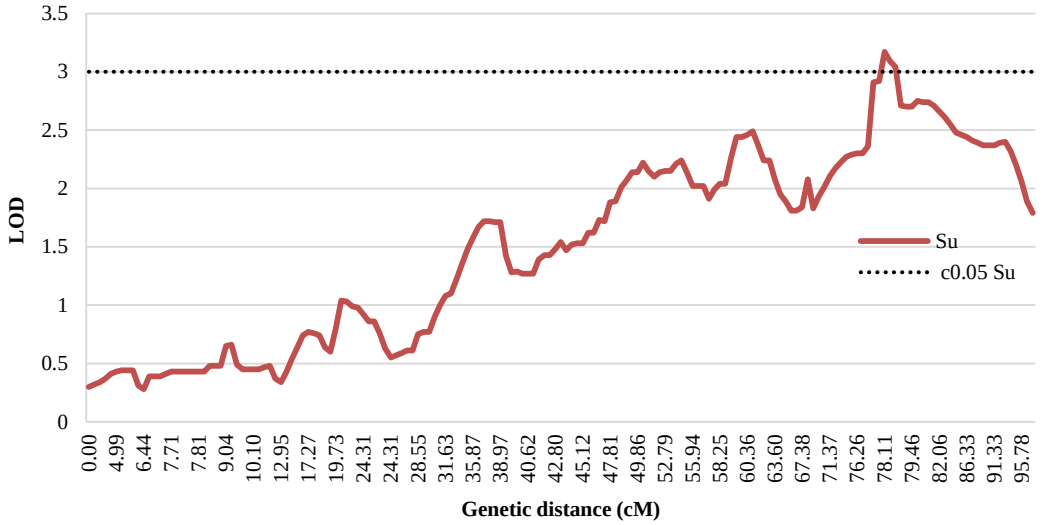


Fig. 3.6 QTL of *qVNN-Su_23.1* detected in LG 23 of Asian seabass. Re for resistance and Su for survival time, c0.05 Su for 0.05 significant level on chromosome-wide.

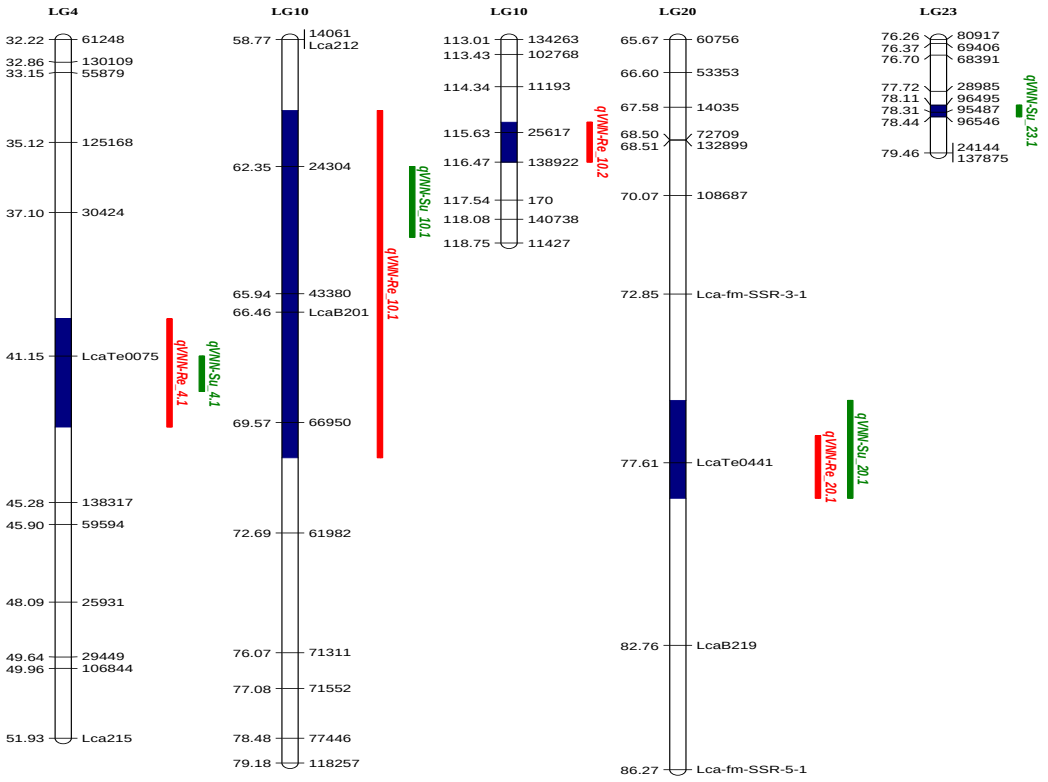


Fig. 3.7 Locations of all the identified QTL for VNN resistance and survival time in the four linkage groups of Asian seabass. Each vertical blank bar represents a linkage group. Each number on the left of linkage group represents a marker's location in cM. Each number or text in the right of linkage group represents marker's name. Each red bar represents the confidence interval of QTL for Resistance. Each green bar represents the confidence interval of QTL for Survival time. Each blue bar represents the confidence interval of QTL for both VNN Resistance and Survival time.

In the present study, multiple loci explaining relatively small to moderate proportions of phenotypic variance for VNN disease resistance in Asian seabass, were mapped. This reflects the polygenetic nature of quantitative traits, which is in line with the classical quantitative genetics theory (Georges 2007). Similar results were also reported in other aquatic species for disease resistance, such as QTL in Atlantic salmon (Gilbey *et al.* 2006), turbot (Rodriguez-Ramilo *et al.* 2011; Rodriguez-Ramilo *et al.* 2013; Rodriguez-Ramilo *et al.* 2014), eastern oyster (Yu & Guo 2006) for resistance against viruses, bacteria and parasites. However, it is also to notice that several studies reported major QTL for resistance to IPN (Houston *et al.* 2008; Moen *et al.* 2009; Houston *et al.* 2010) and

salmonid alphavirus (SAV) (Gonen *et al.* 2015), explaining up to 50% phenotypic variation, in Atlantic salmon. Due to their large effect on the trait, these QTL have already been applied in MAS in Atlantic salmon, greatly reducing the economic losses of the salmon industry (Moen *et al.* 2015). The failure of mapping major QTL in this study could be attributed to a single family being used for mapping, which probably does not have major QTL, or due to no major QTL in Asian seabass population. Multi-family screening could increase the possibility to detect major QTL because a large number of individuals with extreme traits could exist in the mapping populations, thus increasing the detection power for major QTL. In addition, mapping QTL in multi-family would allow precise mapping as the LD block is smaller and cross-validation of these QTL to reduce the false positives. This was demonstrated in Atlantic salmon, as most of these studies were performed on multi-family (Houston *et al.* 2008; Moen *et al.* 2009; Houston *et al.* 2010).Therefore, future studies should focus on QTL mapping on multi-family using GBS, which could allow detection of common QTL for resistance to VNN in Asian seabass. It is worth noting that even if the trait values of parents in the present study were unknown, it would still be feasible to use their offspring to conduct QTL mapping because the assortment and recombination of different alleles in the offspring population could produce a range of phenotypic values (Miles & Wayne 2008).

As the QTL tracked in this study could only explain up to 37% and 37.9% of phenotypic variation for VNN resistance and survival time (Tables 3.2 and 3.3), respectively, the vast majority of missing PVE was not assigned to any QTL. This could be a result of the stringent threshold set by the current QTL mapping approach which could filter out many QTL with very small effects that might explain the missing PVE. GS could overcome this barrier to capture all the QTL with small, moderate and large effect (Hayes & Goddard 2001). GS refers to estimation of genomic breeding values of selected

candidate using genome-wide high-density genetic markers, with an assumption that all the causative QTL are in LD with at least one genetic marker (Hayes & Goddard 2001). Accurate predication the genomic estimated breeding values (GEBV) requires a considerable number of genome-wide markers, preferably SNPs, and genotypes of a large training population (Hayes & Goddard 2001). GBS, possessing the capability to produce tens of thousands of cheap SNPs in a large population, in combination with genotype imputation within the linkage block to reduce the genotyping cost, has the great potential to be applied in aquaculture breeding programs (Sonesson & Meuwissen 2009; Gorjanc *et al.* 2015). With the aforementioned merits, GS is becoming a powerful tool in selective breeding in livestock like dairy cattle, sheep, pig and other domesticated animals, with the expectation of improving their genetic gain in a reduced period in the past decade (Goddard & Hayes 2009). In the vast majority of aquatic species, however, GS is still in its infancy because one of the fundamental features of GS is a large amount of high density genome-wide markers despite the availability of SNP chips for Atlantic salmon (Houston *et al.* 2014) and catfish (Liu *et al.* 2014b). Nonetheless, this obstacle has been removed gradually since sequencing-based genotyping has been rapidly applied in more aquatic species recently. Therefore, future studies could include GS to accelerate the genetic gain for disease resistance in Asian seabass.

3.3.4. Identification of a candidate gene in the significant QTL of *qNNV-Re_20.1* and *qNNV-Su_20.1*

Although fine mapping QTL with a relatively large effect in a narrow region and using the polymorphism in LD with the corresponding QTL are plausible in LD-MAS, LD decays over generations at various degrees and thus compromises the effectiveness of MAS (Georges 2007). Moreover, little information could be presented for understanding of the

mechanism of disease resistance unless the causative polymorphisms underlying gene or regulatory region variance were identified (Georges 2007). Therefore, translation of QTL into corresponding genes containing the quantitative trait nucleotide (QTN) could be essential for comprehensive understanding the mechanism of disease resistance. Thus, to identify candidate genes in the identified QTL region, sequences contacting SSR LcaTe0441 was used as a seed sequence to retrieve 300 kb genomic sequence from Asian seabass genome (kindly provided by Prof. Laszlo Orban of Temasek Life Sciences Laboratory). The transcriptome assembled in Chapter 4 and a previous study (Thevasagayam *et al.* 2015) of Asian seabass ware mapped to the corresponding genomic DNA of *qNNV-Re_20.1* using GMAP (Wu & Watanabe 2005). The result showed that there were 62 mapped genes in this region (Table 3.4). After careful comparison and consideration of the potential functions of the 62 genes, a candidate gene *pcdhαc2*, was proposed to be the corresponding gene controlling *qNNV-Re_20.1*. However, the possibility of other genes being the candidate gene underlying QTL of *qNNV-Re_20.1* and *qNNV-Su_20.1* can not be ruled out. Interestingly, a recent study showed that the corresponding gene underlying a major QTL for resistance to the IPN in Atlantic salmon was the epithelial cadherin (*ecdh*) gene (Moen *et al.* 2015). Ecdh is a calcium-dependent cell-cell adhesion molecule with versatile functions in epithelial cell behavior, tissue formation, cancer suppression, as well as receptor for pathogens (Van Roy & Berx 2008). There was a missense mutation in the coding region of e*cdh*, explaining a majority of phenotypic variation (Moen *et al.* 2015). Further study showed that Ecdh bound to IPN virions, facilitating the internalization of the virus in the susceptible Atlantic salmon individuals while preventing the virus internalization in resistant ones (Moen *et al.* 2015). Surprisingly, *pcdhαc2,* together with *ecdh*, belongs to the cadherin superfamily (Morishita & Yagi 2007).

Table 3.4 Predicated 62 unigenes in QTL *qVNN-Re_20.1* and *qVNN-Su_20.1* in Asian seabass.

Start position	End position	Gene ID	Gene description
3524	4762	comp100618_c0_seq1	gamma-aminobutyric acid receptor subunit beta-2-like isoform X3
3633	6613	comp102955_c0_seq1	gamma-aminobutyric acid receptor subunit beta-4-like isoform X3
3633	6613	comp112774_c0_seq1	gamma-aminobutyric acid receptor subunit beta-4-like isoform X1
22691	28526	comp104236_c0_seq1	immunoglobulin superfamily member 11-like isoform X2
31336	35156	comp118174_c0_seq1	trimethyllysine dioxygenase, mitochondrial-like
46607	47364	comp118073_c0_seq1	55 kDa erythrocyte membrane protein-like isoform 1
50894	59003	comp111512_c0_seq1	vesicle-associated membrane protein 2-like
56267	56421	comp111899_c0_seq1	vesicle-associated membrane protein 3-like
66680	72641	comp103891_c0_seq1	procollagen C-endopeptidase enhancer 1-like
74272	76706	comp116332_c0_seq1	transmembrane protein 88-like
90220	90671	comp100192_c0_seq1	uncharacterized protein LOC100699848
113447	114797	comp114137_c0_seq1	neuroligin-2-like isoform X1
114438	114803	comp102074_c0_seq1	neuroligin-3-like isoform 1
115861	132524	comp105962_c0_seq1	neuroligin-4, X-linked-like
196101	224163	comp107627_c0_seq1	fibroblast growth factor 11-like
215185	223970	comp108047_c0_seq1	fibroblast growth factor 13-like isoform X2
223787	223946	comp115888_c0_seq1	fibroblast growth factor 12-like isoform X2
241701	244312	comp110927_c0_seq1	claudin-7-A-like
258588	268224	comp114045_c0_seq1	Kv channel-interacting protein 1-like isoform X2
260901	261109	comp100688_c0_seq1	Kv channel-interacting protein 2-like isoform X5
313890	315376	comp108700_c0_seq1	uncharacterized threonine-rich GPI-anchored glycoprotein PJ4664.02-like isoform X2
322438	342766	comp100349_c0_seq1	protocadherin alpha-C2-like isoform X1
326222	333568	comp101016_c0_seq1	protocadherin alpha-C2-like
359615	360003	comp120288_c0_seq1	alpha-N-acetylgalactosaminide alpha-2,6-sialyltransferase 1-like
366462	366679	comp114506_c0_seq1	protocadherin-10-like
685171	687551	comp101542_c0_seq1	roundabout homolog 2-like isoform X3
724161	728058	comp102664_c0_seq1	roundabout homolog 1-like isoform X1
778515	778692	comp106378_c0_seq1	dual specificity protein phosphatase 26-like

Continued Table 3.4

812677	819802	comp111020_c0_seq1	protein phosphatase 1D-like isoform X2
812677	819802	comp115610_c0_seq1	protein phosphatase 1D-like isoform X1
840983	842779	comp102668_c0_seq1	T-box transcription factor TBX2b-like
841705	842601	comp121922_c0_seq1	T-box transcription factor TBX2b-like isoform X2
870570	882405	comp114157_c0_seq1	LOW QUALITY PROTEIN acetyl-CoA carboxylase 1
896691	898033	comp111297_c0_seq1	LIM/homeobox protein Lhx1-like
931255	1017849	comp112891_c0_seq1	seizure protein 6 homolog isoform X2
942651	943876	comp103824_c0_seq1	seizure protein 6 homolog isoform X3
1043047	1104197	comp108941_c0_seq1	unconventional myosin-XVIIIa-like isoform X1
1046799	1054297	comp106088_c0_seq1	unconventional myosin-XVIIIa-like isoform X2
1086700	1087892	comp120244_c0_seq1	unconventional myosin-XVIIIa-like isoform X4
1121076	1121732	comp119167_c0_seq1	vascular endothelial zinc finger 1-like
1121684	1151474	comp106090_c0_seq1	vascular endothelial zinc finger 1-like isoform X3
1129704	1129873	comp104531_c0_seq1	uncharacterized protein LOC100698112
1132670	1155774	comp117524_c0_seq1	kinase suppressor of Ras 1-like isoform X1
1151124	1153978	comp118797_c0_seq1	kinase suppressor of Ras 1-like isoform X4
1168509	1168717	comp100167_c0_seq1	oligodendrocyte-myelin glycoprotein-like isoform X2
1179136	1251492	comp117562_c0_seq1	RNA-binding protein Musashi homolog 2-like isoform X2
1179136	1395096	comp109218_c0_seq1	RNA-binding protein Musashi homolog 1-like isoform X1
1179178	1373848	comp121322_c0_seq1	RNA-binding protein Musashi homolog 1-like isoformX1
1348650	1388120	comp110235_c0_seq1	RNA-binding protein Musashi homolog 2-like
1413099	1420985	comp104342_c0_seq1	nuclear fragile X mental retardation-interacting protein 2-like isoform X1
1416673	1420818	comp102552_c0_seq1	nuclear fragile X mental retardation-interacting protein 2-like isoform X2
1428170	1428365	comp100133_c0_seq1	glycine receptor subunit alpha-1-like
1438969	1440907	comp113275_c0_seq1	microtubule-associated protein futsch-like isoform X2
1445850	1454704	comp105529_c0_seq1	GRB2-associated-binding protein 3-like isoform X1
1457123	1462252	comp104216_c0_seq1	APC membrane recruitment protein 1-like
1469590	1471219	comp113498_c0_seq1	probable ribonuclease ZC3H12B isoform X3
1482722	1494229	comp100770_c0_seq1	ezrin-like

Continued Table 3.4

1493123	1493961	comp114850_c0_seq1	moesin-like isoform X3
1531568	1531784	comp108517_c0_seq1	peroxisomal membrane protein 11A-like isoform X2
1560804	1563590	comp101890_c0_seq1	oligophrenin-1-like
1580156	1582678	comp101280_c0_seq1	gap junction beta-1 protein-like isoform X1
1581932	1582600	comp107460_c0_seq1	gap junction beta-1 protein-like isoform X2
1587363	1587666	comp114285_c0_seq1	gap junction alpha-3 protein-like isoform X2
1600115	1600293	comp116137_c0_seq1	TBC1 domain family member 9B-like isoform X1
1600997	1608165	comp116136_c0_seq1	TBC1 domain family member 8B isoform X2
1623120	1627089	comp104106_c0_seq1	neuronal PAS domain-containing protein 2-like isoform X1
1625778	1627050	comp101733_c0_seq1	neuronal PAS domain-containing protein 2-like isoform X2
1676848	1677027	comp108182_c0_seq1	ribosomal RNA processing protein 36 homolog isoform X2
1683207	1683663	comp105941_c0_seq1	uncharacterized protein LOC100690689
1690048	1693564	comp106430_c0_seq1	charged multivesicular body protein 1b-like isoform X2
1696378	1700448	comp105996_c0_seq1	ras-related GTP-binding protein A-like
1703034	1703271	comp114523_c0_seq1	cyclic nucleotide-gated channel rod photoreceptor subunit alpha-like
1703163	1703432	comp121300_c0_seq1	cyclic nucleotide-gated cation channel-like
1710858	1725920	comp117994_c0_seq1	plastin-3
1732103	1741208	comp120629_c0_seq1	transcriptional regulator ATRX-like isoform X1
1732398	1733573	comp110609_c0_seq1	transcriptional regulator ATRX-like
1753579	1753906	comp109943_c0_seq1	coiled-coil domain-containing protein 61-like isoform 2
1758454	1758603	comp119324_c0_seq1	inositol-trisphosphate 3-kinase C-like isoform X2
1769617	1770570	comp111828_c0_seq1	protein phosphatase 1B-like isoform X1
1792648	1797614	comp101920_c0_seq1	poliovirus receptor-related protein 3-like isoform X3
1817525	1822101	comp112361_c0_seq1	protein phosphatase methylesterase 1-like isoform X2
1818191	1819138	comp110475_c0_seq1	protein phosphatase methylesterase 1-like isoform X1
1826714	1832470	comp115056_c0_seq1	mitochondrial uncoupling protein 2-like isoform X2
1832160	1832326	comp115687_c0_seq1	mitochondrial uncoupling protein 2-like
1838966	1843660	comp100725_c0_seq1	ras-related protein Rab-6A-like isoform X2
1842479	1847306	comp112775_c0_seq1	ras-related protein Rab-6A-like isoform X1
1842981	1847278	comp101509_c0_seq1	ras-related protein Rab-6A-like isoform X3

Continued Table 3.4

1852186	1867886	comp113769_c0_seq1	gap junction delta-2 protein-like isoform X1
1867754	1868066	comp102531_c0_seq1	gap junction delta-2 protein-like isoform X2
1897819	1899773	comp116638_c0_seq1	platelet-activating factor acetylhydrolase IB subunit beta-like
1904574	1905828	comp120054_c0_seq1	uncharacterized protein LOC102075665
1907972	1911449	comp115228_c0_seq1	myelin protein zero-like protein 2-like
1922886	1923078	comp121806_c0_seq1	sodium channel subunit beta-4-like
1970154	1974162	comp113144_c0_seq1	Down syndrome cell adhesion molecule-like protein 1-like
1977389	1977667	comp105028_c0_seq1	Down syndrome cell adhesion molecule-like isoform X1
1985050	1999686	comp103373_c0_seq1	rho GTPase-activating protein 35-like

Note: candidate gene protocadherin alpha-C2-like isoform X1 was highlighted by red.

This aroused my speculation that Pcdhαc2 may play a role during the interaction between NNV and Asian seabass. Therefore, the cDNA sequence of *pcdhαc2*, retrieving from the Asian seabass transcrptome (Thevasagayam *et al.* 2015), was firstly examined. This sequence, 7 kb long, contained an ORF of 3 kb long encoding 999 amino acids, and consisting of four exons (Fig. 3.8). Next, the coding sequence (CDS) of *Pcdhαc2* in the two Asian seabass parents was examined, and no SNP in any of the four exons was found, making it impossible to use SNPs to examine the association between this gene and trait. The genomic sequence of *pcdhαc2* in the parents was further examined and a six bp indel and two nucleotides mutations in the 3181 bp locus of the second intron (Fig. 3.8 and 3.9) was found. To conduct association study of *pcdhαc2* to phenotype, a pair of primers was designed for targeting the six bp indel. In addition, an association mapping population from a mass cross was also developed. This association population consisted of 1127 individuals (476 survival and 651 mortalities) from 43 families. Capillary gel electrophoresis of the fluorescence labeled DNA fragments showed that length of PCR product was 254 and 260 bp. The association study of genotypes with phenotypes by Chi-square test showed that the indel of *pcdhαc2* was significantly associated with disease resistance (p = 0.0325). The proportions of individuals in mortality and survival groups with genotype 254_254 were 19.51 and 25.42%, respectively. While for genotype 260_260, the proportions of mortality and survival were 28.11 and 23.32%, respectively (Fig. 3.10). These results could indicate that genotype 254 may be the resistant type, while 260 may be the susceptible type. It could further indicate that the 6 bp deletion is associated with the increased survival proportion. A possible reason could be that indel in the intron could influence the mRNA transcription, splicing variation and further phenotypic expression (Goddard & Hayes 2009). However, identification of the exact cause is impossible with current data. Further study should focus on the functional analysis of the 6 bp indel in *pcdhαc2.*

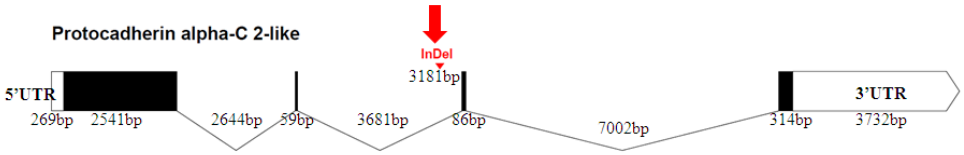


Fig. 3.8 Schematic representation of gene structure of protocadherin alpha-C 2-like in Asian seabass. Each black bar represents an exon. Each number represents the length of a UTR, exon or intron. The six bp indel was located at 3181 bp in the second intron (indicated by the red arrow).

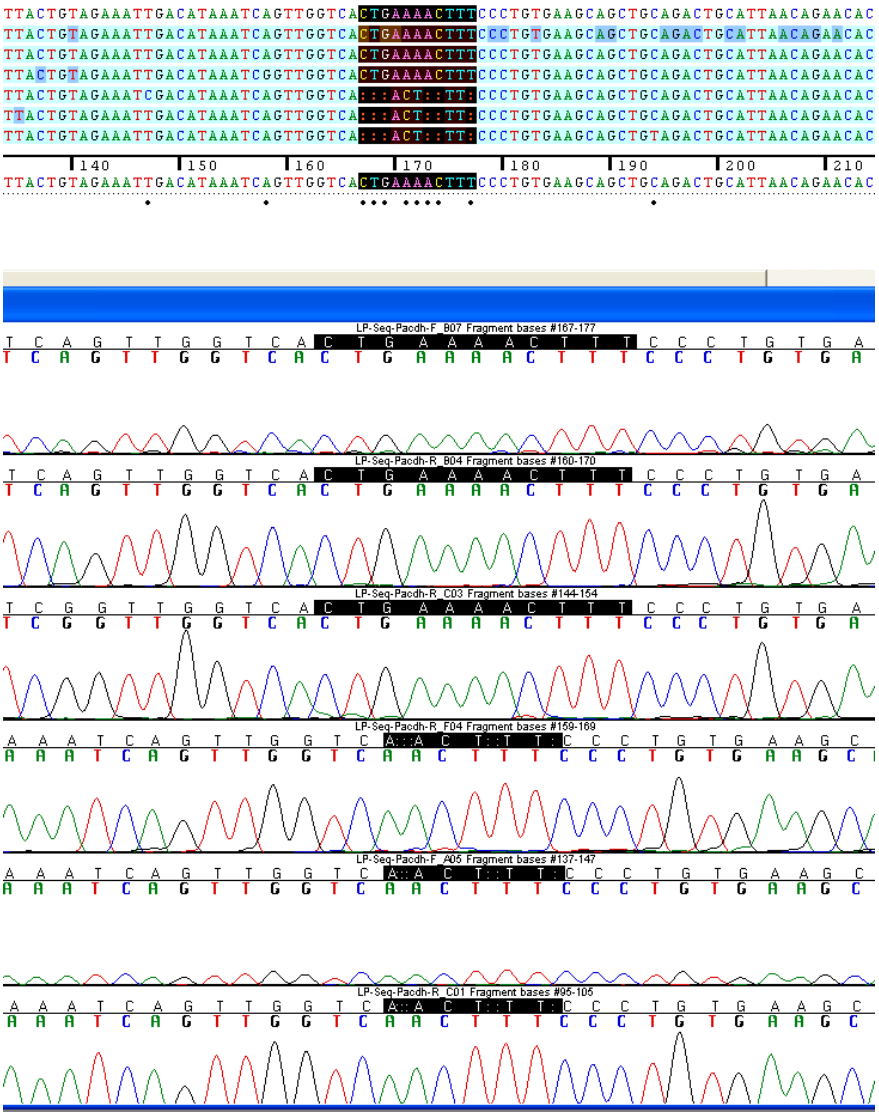


Fig. 3.9 The six bp indel in the second intron of *pcdhαc2* of parents in Asian seabass.

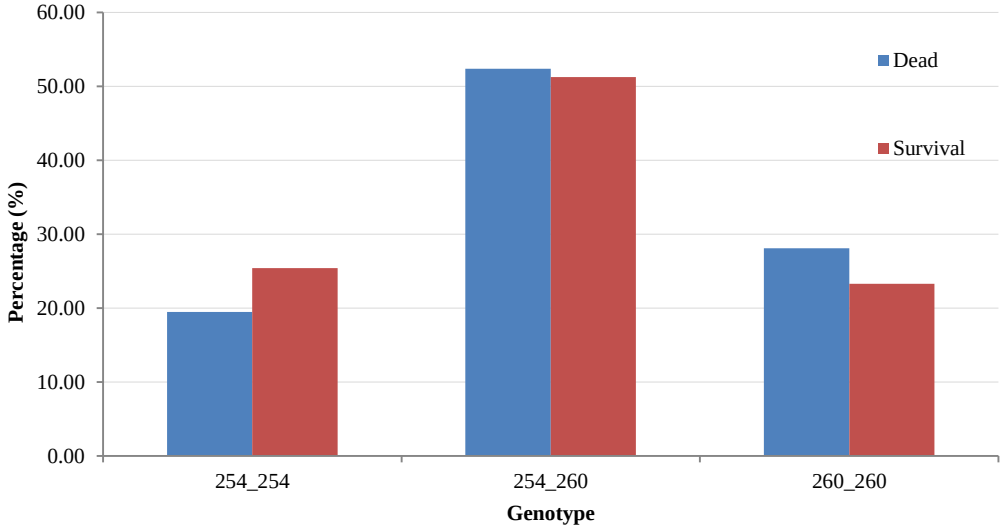


Fig. 3.10 Association between the six bp indel of *pcdhαc2* and VNN disease resistance in the association mapping population of Asian seabass.

The association between the genotype of *pcdhαc2* and phenotype aroused my interest in examining the expression level of this gene in the mock and NNV-challenged fish. qRT-PCR was conducted to determine the expression of *pcdhαc2* in 10 tissues and organs: brain, eye, fin, heart, intestine, kidney, liver, muscle, skin and spleen of NNV-challenged and mock Asian seabass at 5 dpc (Fig. 3.11). The result showed that *pcdhαc2* was significantly induced in the brain (p = 0.0189), muscle (p = 0.0027) and skin (p = 0.0164) after NNV infection, while was suppressed in spleen (p = 0.0013). This indirectly indicates that *pcdhαc2* may play a role in NNV-Asian seabass interaction. Whether the Pcdhαc2 binds to the virion of NNV, thus playing a similar role to Ecdh in Atlantic salmon, or plays a different role, is still unknown. Further studies will focus on elucidating the exact function of Pcdhαc2 in NNV Asian seabass interaction.

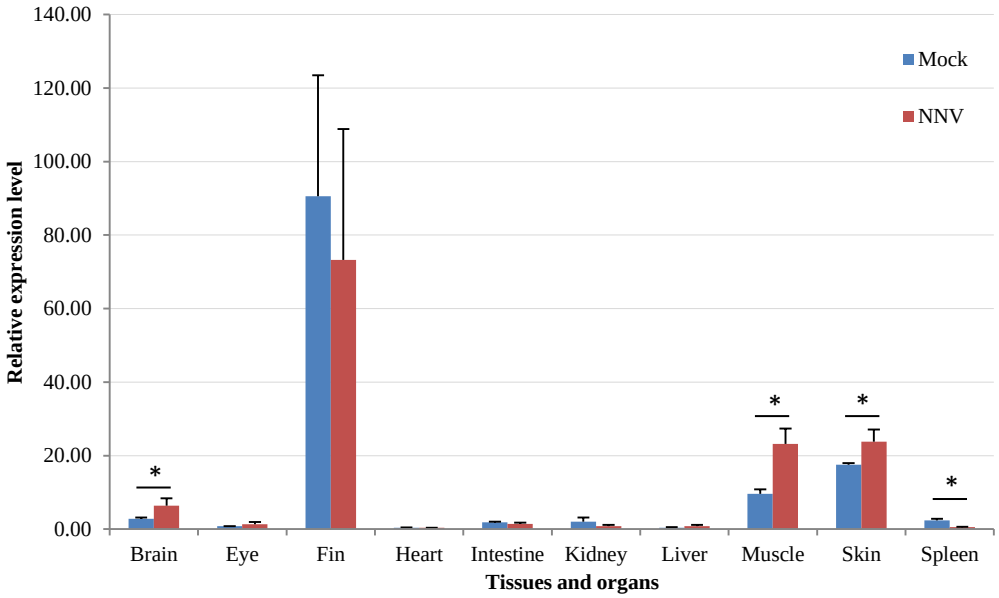


Fig. 3.11 Expression pattern of *pcdhαc2* gene at 5 dpc in mock and NNV-challenged Asian seabass.

3.4. Conclusion

In the present study, a high-density linkage map of Asian seabass, consisting of 3000 genetic markers of SNPs (produced by GBS) and microsatellites, was constructed. Using this map, four moderate QTL for VNN resistance and survival time were mapped. Markers associated these QTL could have the potential to be applied in MAS for VNN resistance in the breeding program of Asian seabass. In addition, a possible candidate gene, *pcdhαc2*, underlying *qNNV-Re_20.1* and *qNNV-Su_20.1*, was identified. It could help us gain an insight into the mechanism of disease resistance in Asian seabass. The indel in *pcdhαc2* could be used as selection marker for VNN resistance breeding in Asian seabass.

Chapter 4. Transcriptome analysis of genes responding to NNV infection in epithelial cells of Asian seabass

4.1. Introduction

With rapid expansion of aquaculture in recent years, fish disease is becoming a major constraint factor challenging the productivity and sustainability of the increasingly important aquaculture industry (Stear *et al.* 2013). Developing effective strategies to combat fish diseases, and thus reducing economic losses caused by them, has become imperative. During the interaction between host and pathogen, fish are able to detect the presence of invading pathogens which triggers the first defense layer through innate immune response (Ellis 2001). Successful defense against pathogen invasion relies heavily on the ability to rapidly activate and effectively mount strong innate immune responses, which leads to elimination of those pathogens in the host (Ellis 2001). Therefore, a very important strategy to reduce disease is to enhance innate immunity in the cultured fish species. Systematic investigation of molecular response such as transcriptome profiling, which could capture nearly complete information about host-pathogen interaction, is essential to understanding of innate immunity. However, until recently, such large-scale and intensive studies of global gene expressions were limited to only model fish species with complete genomic and transcriptomic information available, such as zebrafish (*Danio rerio*) (Hegedűs *et al.* 2009) and medaka (*Oryzias latipes*) (Berger 2010; Wang *et al.* 2015c). With the rapid advancement NGS technology and development of various *de novo* transcriptome assembly algorithms in recent years, it is becoming possible to sequence mRNA (RNA-seq) and assemble millions of reads into large contigs or scaffolds (Ozsolak & Milos 2011). This technology with far-reaching effect has the potential to capture nearly

complete information about transcriptome dynamics under various time-points and conditions of interest, as well as simultaneously identify and quantify gene expressions with high sensitivity and accuracy at genome level (Ozsolak & Milos 2011). Consequently, it could decipher the complicated gene regulation networks and signaling pathways activated and operated in various biological processes (Ozsolak & Milos 2011). With its promising features, high throughput and declining operating cost, this technology has been applied to several aquatic species, including orange-spotted grouper (Huang *et al.* 2011; Lu *et al.* 2012), rainbow trout (Purcell *et al.* 2011), common carp (Li *et al.* 2015), blunt snout bream (Tran *et al.* 2015), half-smooth tongue sole (Zhang *et al.* 2015), and Asian seabass (Xia *et al.* 2013a).Although rapid progresses have been achieved in identifying the transcriptome profiles of several fish species in response to a limited number of diseases, the large gap between comprehensive understanding of the mechanism of immune response and its applications in fish defence against disease still remains to be uncovered.

Epithelial cells, which cover body surface like skin, are extremely important as they form the physical barrier between the outside environment and body interior (Shaykhiev & Bals 2007). Besides providing the barrier, they play a vital role in the initiation, maintenance, regulation and coordination of both innate and adaptive immune response. In innate immunity, they express pattern-recognition receptors (PRRs), which can be activated by pathogen associated molecular patterns (PAMPs) to initiate an innate immune response, including secretion of cytokines and chemokines as well as lysozyme, defensins, collectins and others. In addition, they also help initiate early adaptive immunity by programming responses of dendritic cells to antigen exposure. Furthermore, epithelial cells can directly affect the response of T and B cells including theirs differentiation, proliferation, activation and survival, through chemokines (Schleimer *et al.* 2007; Shaykhiev & Bals 2007). In an aquatic environment, fish constantly and intimately come

into contact with various microorganisms including pathogens, and face an enormous threat from disease infection (Angeles Esteban 2012). Therefore, epithelial tissues, which act as the first line of defense against pathogen invasion, are vital for fish's adaptation in such an environment. A recent study used epithelial cells from fathead minnow (*Pimephales promelas*), challenged with ranavirus, to investigate dozens of genes involved in immune response (Holopainen *et al.* 2012). However, to the best of my knowledge, there is no such study reported in Asian seabass.

The purpose of this study was to identify the DEGs involved in Asian seabass-NNV interaction. The epithelial cell line of Asian seabass was challenged with NNV. mRNA of mock and NNV-infected samples at 6, 12, 24, and 48 hours post-inoculation (hpi) were sequenced and assembled into a transcriptome. The transcriptome consisted of 89026 transcripts, which represented 77750 genes. Subsequently, DEGs and microsatellites were determined. In total, 251 DEGs in response to NNV infection and 24807 SSRs in the transcriptome of Asian seabass were identified. Of these genes, 30 were involved in innate immunity. The top up-regulated genes included *rtp3*, *viperin*, *irf3* and other genes. Two genes, *rtp3* and *viperin*, were selected for further molecular characterization. Furthermore, one SSR in the 3'UTR of *rtp3* was significantly associated with VNN disease in an association study of multiple families of Asian seabass. These results provide valuable information for further analysis and understanding of the molecular interaction between Asian seabass and NNV. The information is also relevant to analysis of microsatellites for genetic studies such as QTL mapping in Asian seabass.

4.2. Materials and methods

4.2.1. NNV and Asian seabass epithelial cell line

NNV was prepared as described in Chapter 2. Asian seabass epithelial cells were maintained as described in Chapter 2.

4.2.2. Challenge Asian seabass epithelial cells with NNV

To challenge SB cells with NNV, about 1.4×10^7 cells were seeded in a 75 cm^2 flask (BD biosciences, Franklin lakes, USA) and 90% cell confluence was reached after 4 hours incubation at 28°C. NNV at multiplicity of infection (MOI) of 10 was added to the cell culture to form the NNV challenged samples. Used L-15 medium, equivalent in volume to NNV challenged sample, was added to a separate flask of cell culture as mock. Every treatment included three biological replicates.

4.2.3. Total RNA extraction, assessment and pooled sample processing

At 6, 12, 24 and 48 hpi, mock and NNV challenged Asian seabass SB cells were harvested and total RNA was extracted using Trizol (Life Technologies, Carlsbad, USA) following the manufacturer's instructions. A total of 24 total RNA samples, comprised of three biological replicates for each of the four time points of the two treatments, were obtained. The whole process followed the procedure described in Chapter 2, except for several steps prior to RNA extraction. These steps were: washing the monolayer cells with PBS for three times before adding 3 ml of Trizol onto the cell monolayer and incubating for 5 min to lysate the cells. To primarily assess RNA quality, 1 µl of RNA was electrophoresed on 1% agarose gel stained with 1% EB and visualized using Gel DOC XR (Bio-Rad Laboratories, Hercules, USA). Primary RNA concentration was determined by Nanodrop (Thermo Fisher Scientific, Waltham, USA). To further assess the RNA quality and quantity, total RNA was submitted to Agilant 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA), as it produced an assessment with higher accuracy. Before assessment

in Agilant 2100 Bioanalyzer, all the RNA samples were diluted with an appropriate number of times, according to the primary results determined by Nanodrop, to a final concentration ranging from 50 to 500 ng/µl. Agilent RNA 6000 Nano Kit and RNA Nano Chip were used and manufacturer's instructions were followed. Briefly, a new RNA chip was put on the chip priming station before adding 9 µl of gel-dye mix in the right well. The plunger was moved to the position of 1 ml following the closing of chip priming station. The plunger was pressed until it was held by the clip and waited for exactly 30 s before releasing the clip and waiting for 5 s. Plunger was slowly pulled back to the 1 ml position before releasing chip from chip priming station. Another two wells were added with 9 µl of gel-dye mixture followed by adding 5 µl of RNA marker to the rest of wells. The ladder well was added 1 µl of prepared ladder and the sample wells were added 1 µl of RNA samples. The chip was vortexed at 2400 *x* g for 1 min before submitting to Agilant 2100 Bioanalyzer. After quality and quantity assessment by Agilant 2100 Bioanalyzer, only these RNA samples with integrity number above 7.5 were kept for further analysis. According to quantity assessment from Agilant 2100 Bioanalyzer, equal amounts of RNA from each treatment at each time-point were pooled before being submitted to Macrogen, In, South Korea for RNA-seq.

4.2.4. Construction of sequencing libraries and Illumina PE sequencing

Eight sequencing libraries from the pooled mock and NNV-challenged RNA samples representing four time-points were constructed. One µg of total RNA for each sample was used for library construction using Tureseq^{TM} RNA sample prep Kit (Illumina, San Diego, USA). Briefly, total RNA was treated with DNase to remove the genomic DNA contamination, and mRNA with poly (A) was enriched using magnetic beads coated with oligo dT subsequently. The collected mRNA was then fragmented and first-strand cDNA

was reverse-transcribed with random hexamer-primers, after which, second-strand cDNA was synthesized and purified. The purified double-strand cDNA was then repaired, tailed and added with sequencing index and adapters. Proper DNA fragments were selected and enriched for creation of final sequencing libraries. The libraries were subsequently sequenced by Illumina HiSeq™ 2000 (Illumina, San Diego, USA) to produce 2 × 101 bp sequence reads.

4.2.5. Processing of sequence reads

Raw reads were processed by NGS QC toolkit (Patel & Jain 2012) to remove adapters, low quality reads (Q < 20) and unpaired reads using command of 'perl IlluQC.pl -pe f1.fq r1.fq 2 A'. The output included GC content distribution, average quality distribution, base composition and percentage of reads for different quality score ranges at each base position.

4.2.6. Transcriptome *de novo* assembly

After filtration, only high-quality reads were kept and subjected to the transcriptome *de novo* assembly platform Trinity (version 20140717) (Grabherr *et al.* 2011; Haas *et al.* 2013) for transcriptome assembly. Basically, there are three stages, namely Inchworm, Chrysalis and Butterfly, for transcriptome assembly in Trinity. Inchworm is responsible for greedy assembly of initial contigs that represent genes end clusters of similar genes. Chrysalis is responsible for generating de Bruijn transcript graphs for each cluster and partitions reads between clusters. Butterfly is used to resolve transcripts, with alternative splicing and paralogy, independent for each cluster. The following commands and parameters in Trinity platform were used for Asian seabass transcriptome *de novo* assembly.

For *de novo* assembly:

'perl trinityrnaseq_r20140717/Trinity --seqType fq --left all_ left.fastq --right all_ right.fastq --output Trinity_OUT_200/ --CPU 24 --bflyHeapSpaceMax 40G --bflyCPU 24 --JM 250G --full_cleanup --min_kmer_cov 5 --inchworm_cpu 24 --min_glue 10 --min_per_id_same_path 90'.

For calculation of statistics about the transcriptome:

'perl trinityrnaseq_r20140717/util/TrinityStats.pl Trinity_OUT_200.fasta > Trinity.fasta_stats.txt'.

4.2.7. Identification of differentially expressed transcripts

After obtaining the *de novo* assembled transcriptome by Trinity, the abundance of each transcript was estimated by aligning each read back to the transcriptome using the Trinity package RSEM. The number of reads mapped to each transcript was counted for each sample and then normalized as value of fragments per kilobase of transcripts per million fragments mapped (FPKM). The false discovery rate (FDR) was used to determine the threshold p value for multiple tests. The FPKM values of the samples were compared using the R package edgeR provided by Trinity. Transcripts were considered as DEGs if their FDR values were less than 0.05 (significance level) and their FPKM values were more than four between the matched time points (4 fold change). The following commands and parameters were used for identification of DEGs.

For reference preparation of transcriptome:

'perl trinityrnaseq_r20140717/util/align_and_estimate_abundance.pl --thread_count 24 --est_method RSEM --aln_method bowtie2 --trinity_mode --transcripts Trinity_OUT_200.fasta --prep_reference'

For reads alignment to reference and reads calculation on each sample:

'perl trinityrnaseq_r20140717/util/align_and_estimate_abundance.pl --thread_count 24 --est_method RSEM --aln_method bowtie2 --trinity_mode --transcripts Trinity_OUT_200.fasta --seqType fq --left each_sample_1.fastq_filtered --right each_sample_2.fastq_filtered --output_prefix each_sample'

4.2.8. Gene annotation and ontology of DEGs

The sequence of identified DEGs were submitted to Blast2GO (Conesa *et al.* 2005) for blast, annotation, mapping and gene ontology, with default parameters.

4.2.9. qRT-PCR for expression patterns of *viperin* and *rtp3* in Asian seabass after NNV infection

The cDNA from tissues and organs of three-month old Asian seabass challenged with NNV (Chapter 3) were used to determine the expression patterns of *viperin* and *rtp3* using qRT-PCR as described in Chapter 3. The gene sequences of *viperin* and *rtp3* were retrieved from the assembled Asian seabass and were used to design primers. Two primer pairs targeting ORFs of *viperin* (Lca-viperin-q) and *rtp3* (Lca-RTP3-2-q) were designed (Appendix Table A3). Data analysis followed the method as described in Chapter 3.

4.2.10. Cloning, transformation and cellular localization of fusion protein Viperin-EGFP and RTP3-EGFP

The gene sequences of *rtp3* and *viperin* were retrieved from the assembled Asian seabass transcriptome. The ORFs of *viperin* and *rtp3* were identified using EditSeq in DNASTAR Lasergene 11 (DNASTAR, Madison, USA). Primer pairs with sequence of HindIII restriction site were synthesized and used to amplify the ORFs without stop codon of *viperin* (Lca-viperin-Hindiii) and *rtp3* (Lca-RTP3-2-Hindiii), with the purpose of

incorporating the cutting sites for cloning (Appendix Table A4). The PCRs were performed using the same procedures as described in Chapter 2. PCR products were separated on a 0.8% agarose gel, and the DNA bands corresponding to the correct sizes were excised and purified by a QIAquick gel purification kit (Qiagen, Hilden, Germany). These DNA were digested with HindIII (New England Biolabs, Ipswich, USA) and ligated to pEGFP-N1 vector (Clontech Laboratories, Japan) to create constructs of Viperin-pEGFP and RTP3-pEGFP. The constructs were transformed into *Escherichia coli* DH5α. The construct DNA were extracted using Plasmid DNA Purification Kit (Qiagen, Hilden, Germany) and verified by DNA sequencing. The correct constructs were extracted using EndoFree Plasmid Kits (Qiagen, Hilden, Germany) for transfection. Before transfection, Asian seabass epithelial cells were grown to a confluence of 80% on a glass bottom petri dish (MatTek Corporation, Ashland, USA). The transfection of the two constructs was performed on these cells using TurboFect Transfection Reagents (Thermo Fisher Scientific, Waltham, USA) following the manufactures' instructions. At 48 hours post transfection, cells were washed with 1x PBS for three times before staining with DAPI. The cells were observed and signals of GFP and DAPI were detected on the Zeiss LSM 510 Meta Confocal Microscope (Carl Zeiss, Oberkochen, Germany).

4.2.11. Association of *viperin* and *rtp3* with VNN resistance in multiple families of Asian seabass challenged with NVN

The association population was the same population as described in Chapter 3. Primer pairs targeting the microsatellites in *viperin* (Lca-viperin-SSR1) and *rtp3* (Lca-RTP3-SSR1 and Lca-RTP3-SSR2) (Appendix Table A4) were used to genotype the association population following the same procedures as described in Chapter 2.

4.2.12. qRT-PCR for validation of expression of DEGs in RNA-seq data

qRT-PCR was conducted on Biorad iQ5 (Bio-Rad Laboratories, Hercules, USA), using SYBR Green as fluorescent dye, to validate the RNA-seq data. A total of 19 pairs of primers targeting 19 (15 up-regulated and 4 down-regulated) genes were designed (Appendix Table A3). A total of 24 RNA samples, comprised of three biological replicates for each time point of 6, 12, 24 and 48 hpi for the mock and NNV-challenged groups, were used for qRT-PCR. Removing DNA contamination, converting RNA to cDNA and qRT-PCR were conducted using the same procedures as described in Chapter 2. Data were analyzed by the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen 2001) for relative quantifications and fold change of each gene was determined using *EF1a1* as a reference gene (Wang *et al.* 2014d) as described in Chapter 2.

4.2.13. Mining microsatellites in the transcriptome

To avoid redundancy, only longest isoform of each gene was extracted from the assembled transcriptome and subsequently used for microsatellite mining. These transcripts were subjected to the SSR identification tool GMATo version 1.0 (Wang *et al.* 2013) and microsatellites with motifs from two to six bp in size repeating at least five times were determined. The following commands and parameters were used for identification of SSRs in the Asian seabass transcriptome:

'perl gmat.pl -r 5 -m 2 -x 6 -s 0 -i longest_transcritps.fasta'

4.2.14. Data deposition

All the raw reads were submitted to the sequence reads archive (SRA), NCBI database with an accession number of PRJNA283461.

4.3. Results and discussion

4.3.1. Processing of raw sequencing reads

The eight libraries representing eight pooled RNA samples of mock and NNV challenged cells were sequenced in a single lane by the Illumina Hiseq 2000 platform, generating over 415 million pair-end raw reads of 2 x 101 bp. After processing by NGS QC Toolkit (Patel & Jain 2012) for removing adopters, low quality (Q < 20) and single paired-end reads, about 385 million (92.72%) remained as clean reads (Table 4.1). The percentage of clean reads of each sample varied from 92.30 to 92.95% (Table 4.1). This is the typical output produced by Illumina Hiseq 2000.

Table 4.1 Statistics of sequence reads from eight libraries of Asian seabass.

Sample	Total bases	Read count	Q20 (%)	Clean reads	Percentage (%)
mm6h	5123967754	50732354	95.83	47154842	92.94826335
mm12h	5276012346	52237746	95.81	48545902	92.93261237
mm24h	5325588600	52728600	95.63	48832016	92.61011292
mm48h	4749494296	47024696	95.68	43571948	92.65758571
mv6h	5699679672	56432472	95.85	52447378	92.93829623
mv12h	4875829742	48275542	95.72	44787222	92.77414638
mv24h	5351955660	52989660	95.62	49058078	92.58047325
mv48h	5505438088	54509288	95.49	50312172	92.30018194
total	41907966158	414930358		384709558	92.71665729

Note: Q: phred quality scores for reads

4.3.2. Transcriptome *de novo* assembly

To obtain a transcriptome of Asian seabass, all the clean reads were submitted to the Trinity platform to *de novo* assemble the transcriptome of Asian seabass. The assembled transcriptome consisted of 89026 transcripts, representing 77750 genes (Table 4.2). It is worth to note that the number of gene in this study is much higher than the typical number of gene of fish. The obvious reason is the *de novo* assembled method, which

apparently produces more 'genes' because some genes could be broken apart. The total length of the transcriptome was 105 Mb and the average length of the transcripts was 1174.9 bp (Table 4.2). The number of transcripts is highly consistent with a previous study, which reported 83911 transcripts obtained from the intestine of Asian seabass using Roche 454 technology (Xia *et al.* 2013a). However, the average length found in that study was much shorter, only 747 bp (Xia *et al.* 2013b). This could be due to differences in sampling and sequencing platforms, as well as assembly algorithms. The number and average length of transcripts are fairly close to those of the transcriptome of crimson spotted rainbowfish (*Melanotaenia duboulay*) (107749 transcripts, 960.63 bp) sequenced by the Illumina Hiseq 2000 platform (Smith *et al.* 2013). In contrast, the numbers of transcripts in the transcriptomes of common carp, blunt snout bream and orange-spotted grouper were 130292 (Li *et al.* 2015), 253439 (Tran *et al.* 2015) and 204517 (Lu *et al.* 2012), respectively, much higher than that in Asian seabass. The average length of transcripts for common carp (1400.57 bp) (Li *et al.* 2015) was slightly longer than that in Asian seabass. In contrast, those of blunt snout bream (998.03 bp) (Tran *et al.* 2015) and orange-spotted grouper (527 bp) (Lu *et al.* 2012) were shorter than that in Asian seabass. The differences could be due to number of samples, libraries and assembled platforms, as well as the species used for study. Nevertheless, this demonstrates that this study generated a better assembled transcriptome. The length distribution of the Asian seabass transcriptome is presented in Fig. 4.1. The lengths of the majority transcripts (60630, 68.10%) were concentrated in a range of 200 to 999 bp, followed by 12073 (13.56%) in 1000 to 1999 bp, 6612 (7.43%) in 2000 to 2999 bp, 4160 (4.67%) in 3000 to 3999 bp, 2454 (2.76%) in 4000 to 4999 bp and 3097 (3.48%) for more than 5000 bp. Those results imply that Illumina sequencing has the potential to rapidly capture a larger number of transcripts, and the Trinity platform has the ability to effectively assemble the transcriptomes of non-model organisms without solid reference genomes.

Table 4.2 Statistics of transcriptome assembled by Trinity in Asian seabass.

Item	Number
Total trinity genes	77750
Total trinity transcripts	89026
GC content	45.70%
Contig N50	2617 bp
Median contig length	495 bp
Average contig length	1174.9 bp
Total assembled bases	104593557 bp

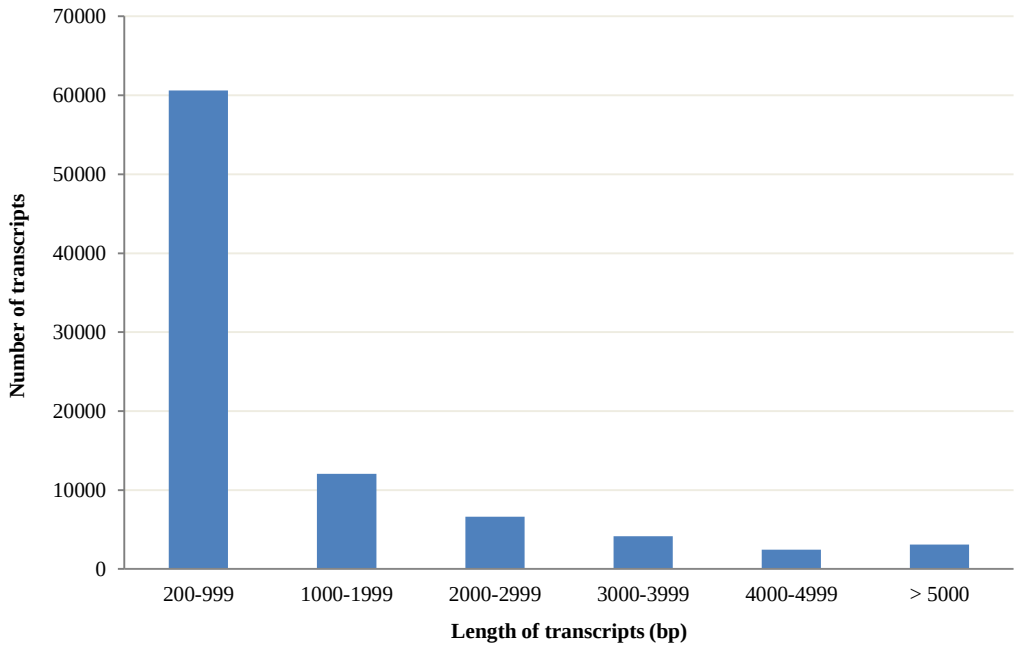


Fig. 4.1 Length distribution of contigs from *de novo* assembled transcriptome of Asian seabass.

4.3.3. Identification, gene annotation and gene ontology of DEGs

The outcome of host virus interaction depends on how quickly and strongly the host mounts an immune response, as well as how fast the virus replicates itself using the virulence factor to compromise host immunity before and even at the time when enough immune response is mounted by the host (Workenhe *et al.* 2010). In this study, differentially expressed transcripts were analyzed by RSEM and edgeR in the Trinity

platform. As shown in Fig. 4.2, the number of differentially expressed transcripts at 6 hpi and 12 hpi, were 17 and 46, respectively. However, just at 24 hpi, the number increased to 274, about 6 times more than that at 12 hpi, and the number continued to climb to 428 at 48 hpi. This could suggest that most of the differentially expressed transcripts are late response genes. This could also indicate that NNV replication in the cells was not checked and suppressed before the development of cell-mediated outcome, resulting in massive cellular damages. Compared with the total time-points from other similar studies such as orange-spotted grouper challenged with NNV (6 and 33 hpi) (Lu *et al.* 2012) and iridovirus (48 hpi) (Huang *et al.* 2011), and blunt snout bream challenged with *A. hydrophila* (4, 12, 24 hpi) (Tran *et al.* 2015), this study covered a wider range time-points. This enables us to gain more details about the fluctuations of gene expression, and thus achieve a more complete understanding of the molecular response of Asian seabass to NNV invasion.

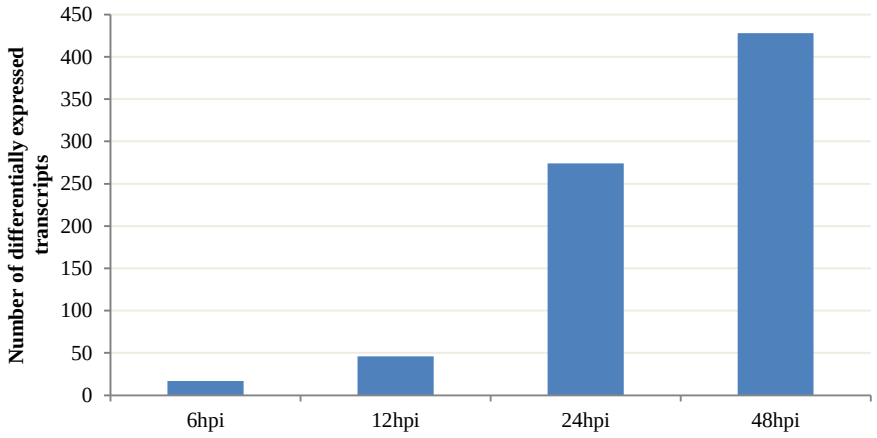


Fig. 4.2 Number of differentially expressed transcripts in four time-points after NNV infection in Asian seabass.

To annotate the differentially expressed transcripts, all the 428 transcripts were submitted to Blast2GO (Conesa *et al.* 2005). After gene annotation by blastx against the protein database of non-redundant protein sequences (nr), 318 (74.30%) of 428 transcripts with an average length of 2152 bp were annotated, leaving 110 (25.70%) un-annotated

transcripts with lengths ranging from 317 to 13659 bp and average length of 1183.04 bp (Table 4.3). The possible reasons could be due to, 1) some of these transcripts are novel or Asian seabass specific, 2) sequencing errors and spurious sequences introduced by indirect RNA sequencing during various steps of RNA-seq (Ozsolak & Milos 2011), and 3) mis-assembled contigs or artefacts by assembly algorithm.

Table 4.3 Summary of unigenes and un-annotated genes in the transcriptome of Asian seabass.

Item	Number	Length (bp)			Fold change	
		Average	Max.	Mini.	Average	Max.
unigene	251	2270.57	7117	359	-	-
up-regulated unigene	247	2276.36	7117	359	60.31	9118.63*
down-regulated unigene	4	1912.75	3353	896	10.21	18.98
un-annotated gene	110	1883.04	13659	317	-	-
up-regulated un-annotated gene	102	1970.5	13659	317	44.01	1043.67*
down-regulated un-annotated gene	8	767.88	1503	364	15.54	21.39

Note: - not calculated, * calculation on data of 48 hpi

There could be redundancies for the protein coding genes in the annotated transcripts. Therefore, to identify unigenes, for gene with the same name, only one was retained and duplicates were removed from the Blast2GO output. The results showed that 251 (78.93%) of 318 transcripts were identified as unigenes with lengths ranging from 359 to 7117 bp and an average length of 2271 bp. Of these 251 unigenes, 247 (98.41%) were identified as up-regulated genes with an average fold-change of 60.31 and average length of 2277 bp, and four (1.59%) were down-regulated genes with an average fold-change of 10.21 and average length of 1913 bp (Table 4.3). This indicates that epithelial cells could sensitively regulate expressions of related genes by up-regulating genes which have negative effects on the virus and down-regulating genes which have positive effects on the virus. In turn, this could result in mobilizing new resources to restrict virus infection. Alternatively, viruses could deliver virulence effectors to effectively modify the host gene

expression and exploit the host metabolism machinery for virus replication and invasion of new cells.

Further analysis of the blastx results found that, among the top most-hit species, more than 400 hits were four cichlid fish species: *Oreochromis niloticus* with 609 hits, *Maylandia zebra* with 530, *Haplochromis burtoni* with 467 and *Pundamilia nyererei* with 415. Other species included *Stegastes partitus* with 474, *Poecilia reticulate* with 467 and *Poecilia formosa* with 465 (Table 4.4).This may indicate that Asian seabass is genetically more related to cichlid fish than other species. Alternatively, this could be due to cichlid fish having more information than other species in the protein database. This is more likely the case, since a recent genomic syntenic study using thousands of sequence tags showed that Asian seabass was closer to European seabass than nile tilapia (Wang *et al.* 2015a), despite the latter being the most-hit species in the blastx results.

Table 4.4 Most-hit-species after blasting of differentially expressed transcripts.

Species	**Blast hits**
Oreochromis niloticus	609
Maylandia zebra	530
Stegastes partitus	474
Poecilia reticulata	467
Haplochromis burtoni	467
Poecilia formosa	465
Pundamilia nyererei	415
Larimichthys crocea	397
Neolamprologus brichardi	384
Danio rerio	334
Esox lucius	316
Cynoglossus semilaevis	284
Oryzias latipes	282
Oncorhynchus mykiss	281
Notothenia coriiceps	254
Xiphophorus maculatus	248
Takifugu rubripes	212
Astyanax mexicanus	155
Tetraodon nigroviridis	133
Lepisosteus oculatus	103
others	1238

Gene ontology (GO) is a widely used representation to classify gene functions and their products in organisms. Based on the GO analysis of genes in this study, a total of 70 (27.89%) transcripts were assigned to 146 GO terms, covering three GO domains, with an average of two terms per gene. Of these, 65 genes were assigned to biological process (BP), 59 to molecular function (MF) and 22 to cellular component (CC). The low rate of GO annotation for transcripts of Asian seabass could be due to the uninformative description of gene functions in the protein database. This result is consistent with the number of genes with successful GO annotation in common carp (Li *et al.* 2015). For BP, the GO terms with the most assigned genes were: cellular process (GO:0009987) with 78 transcripts, and

single-organism process (GO:0044699) with 76 genes; the other GO terms were biological regulation (GO:0065007) with 46 transcripts, response to stimulus (GO:0050896) with 45, and metabolic process (GO:0008152) with 44 (Fig. 4.3). For MF, a large number of transcripts, 78, were assigned to binding (GO:0005488), 47 to catalytic activity (GO:0003824), and 20 to transporter activity (GO:0005215) (Fig. 3). For CC, most transcripts, 57 of them, performed their functions in cell (GO:0005623), 55 in membrane (GO:0016020), 25 in organelle (GO:0043226) (Fig. 4.3). These results indicate that the identified DEGs of Asian seabass in response to NNV infection had diverse functions, and were involved in various processes and pathways in different parts of cells. It further implies that the NNV infection produced multi-dimensional and multi-factorial impacts on Asian seabass cells, and that the immune response is a very complex process.

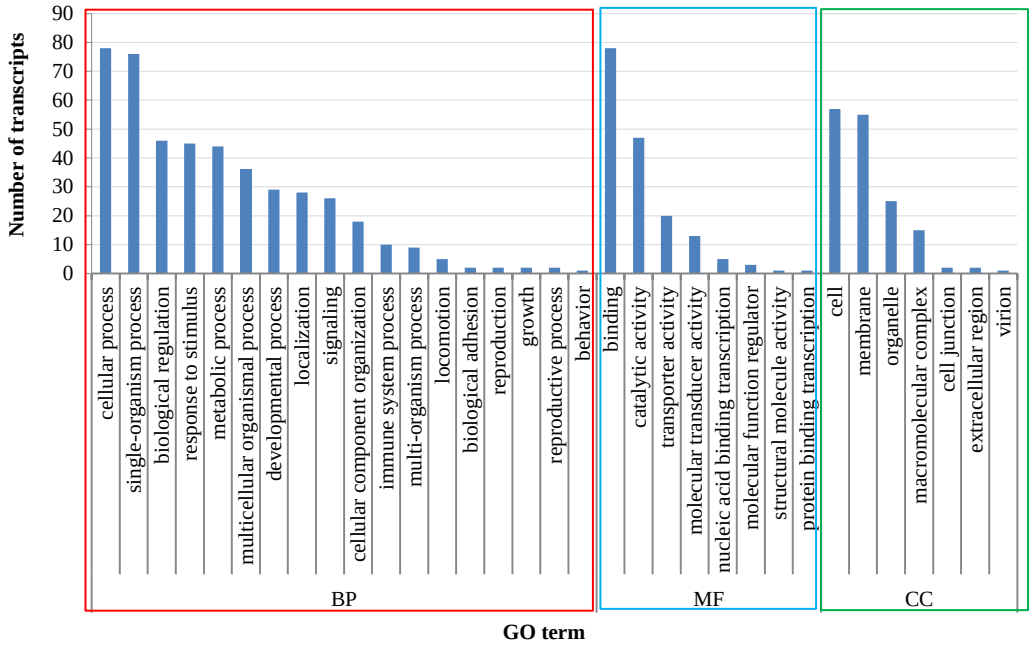


Fig. 4.3 GO distribution of differentially expressed transcripts in Asian seabass after NNV infection. BP: bilogical process, MF: moleuclar function, CC: cellular component.

4.3.4. DEGs of Asian seabass involved in innate immunity after NNV infection

Innate immunity plays a critical role in defending against pathogen infection. Fish immune responses have not been as well characterized as their mammalian counterparts. Nevertheless, they do share many features of immune responses, as many key homologous genes have been identified in several fish species (Reyes-Cerpa *et al.* 2012). Fish epithelial cells, forming physical barriers between body and external environment, play an essential role in the complex immune system (Vasta *et al.* 2011; Holopainen *et al.* 2012; Reyes-Cerpa *et al.* 2012). Viruses are unable to multiply on their own and their replication is dependent on the host's cellular machinery. They need to enter the host cells and gain access to the synthetic machinery. Once inside the cell, virus associated molecular patterns (VAMPs) like ssRNA and dsRNA are recognized by the host's PRRs like the toll-like receptors (TLRs). The interaction of VAMPs and PRRs could initiate the recruitment of adapters and downstream transcription factors to produce a variety of cytokines including inflammatory cytokines and chemokines, as well as IFN which in turn produces ISGs (Workenhe *et al.* 2010).

In this study, a suite of genes, in terms of both number and magnitude, were identified from an epithelial cell line of Asian seabass to be involved in innate immunity, which could enable the infected and neighboring cells to limit the spread of viral pathogens (Ivashkiv & Donlin 2014). These genes, from a variety of cytokines including pro-inflammatory cytokines and interferon (IFN) chemokines, were strongly up-regulated after NNV infection in the epithelial cells (Table 4.5).

Table 4.5 30 differentially expressed genes involved in innate immunity of Asian seabass after NNV infection.

Gene id	P Value	FDR	Gene expression level (FPKM)								Fold change	Sequence description
			mm 6h	mm 12h	mm 24h	mm 48h	mv 6h	mv 12h	mv 24h	mv 48h		
c35231_g1_i1	1.64E-29	2.42E-25	0.13	0.12	0.23	0.05	0.07	1.13	24.85	474.17	9118.63	protein asteroid homolog 1-like
c3737_g1_i1	2.24E-21	1.50E-17	0.03	0.02	0.00	0.01	0.04	0.08	1.33	19.17	1916.6	interferon-induced very large gtpase 1-like
c11620_g1_i1	5.05E-21	2.86E-17	0.13	0.13	0.27	0.25	0.53	0.84	46.45	256.58	1022.23	heat shock protein 30-like
c11709_g1_i1	5.40E-22	4.42E-18	0.26	0.17	0.36	0.35	0.83	4.75	28.75	262.54	758.79	receptor-transporting protein 3-like
c12370_g3_i1	1.48E-21	1.09E-17	3.61	2.75	2.49	3.64	5.09	22.39	407.74	1722.26	473.67	heat shock protein 70
c10062_g1_i2	5.38E-13	9.45E-10	0.00	0.00	0.00	0.09	0.09	0.09	1.17	28.72	305.52	grass carp reovirus -induced gene 2i
c144_g1_i1	1.18E-15	3.23E-12	0.13	0.27	0.14	0.25	0.61	2.57	10.59	74.13	295.35	C-X-C motif chemokine 6-like
c51268_g1_i1	8.23E-11	9.20E-08	0.03	0.09	0.00	0.04	0.11	0.04	0.50	6.70	159.48	interleukin-22 receptor subunit alpha-2-like isoform x2
c12542_g9_i1	1.73E-14	3.75E-11	0.06	0.05	0.22	0.16	0.25	2.68	6.04	21.53	137.11	interferon regulatory factor 3
c65346_g1_i1	6.02E-11	6.93E-08	0.15	0.07	0.23	0.05	0.18	0.13	1.10	6.82	131.13	interferon-induced protein with tetratricopeptide repeats 5-like
c27937_g1_i1	3.62E-08	1.64E-05	0.00	0.00	0.04	0.04	0.05	0.17	0.80	4.00	95.12	suppressor of cytokine signaling 1-like
c8372_g1_i1	2.09E-13	3.86E-10	0.23	0.26	0.53	0.75	0.25	0.34	4.65	55.70	73.88	interferon inducible mx protein
c27550_g1_i1	2.52E-12	3.79E-09	0.24	0.12	0.46	0.28	0.25	0.17	1.35	18.24	64.45	probable E3 ubiquitin-protein ligase herc6-like isoform x2
c44647_g1_i1	7.42E-07	2.04E-04	0.00	0.22	0.08	0.08	0.48	0.24	0.69	5.32	63.29	interferon alpha 1
c9285_g1_i1	9.08E-12	1.20E-08	0.08	0.07	0.23	0.25	0.16	0.31	1.54	14.61	58.2	probable atp-dependent rna helicase dhx58
c12674_g1_i1	7.14E-11	8.09E-08	0.10	0.10	0.49	0.65	0.16	0.10	3.56	27.18	41.81	interferon- double-stranded rna-activated protein kinase-like
c57397_g1_i1	6.76E-10	5.86E-07	0.50	0.27	0.64	0.49	0.26	0.26	1.74	17.65	35.88	tripartite motif-containing protein 16-like
c5512_g2_i1	1.60E-10	1.64E-07	1.19	1.32	12.03	16.95	3.60	4.49	44.75	565.28	33.35	interferon stimulated protein 15
c16863_g1_i1	1.36E-07	5.09E-05	0.14	0.04	0.26	0.23	0.13	0.08	0.66	6.79	29.39	stimulator of interferon genes
c42741_g1_i1	9.85E-08	3.86E-05	0.08	0.18	0.12	0.15	0.42	1.84	4.27	3.81	25.94	E3 ubiquitin-protein ligase trim39-like
c7898_g1_i1	6.73E-20	3.04E-16	0.00	0.04	0.04	0.00	0.00	0.04	1.96	56.81		viperin
c591_g1_i1	6.52E-18	2.19E-14	0.00	0.00	0.00	0.00	0.05	0.16	1.18	21.32		probable E3 ubiquitin-protein ligase herc4-like
c38898_g1_i1	9.22E-16	2.61E-12	0.00	0.00	0.00	0.00	0.15	0.23	3.21	36.69		CC chemokine
c10062_g1_i1	5.51E-14	1.13E-10	0.00	0.00	0.00	0.00	0.00	0.32	2.11	34.52		gig2-like protein
c9740_g1_i2	8.59E-14	1.67E-10	0.00	0.09	0.07	0.00	0.04	0.43	1.29	10.82		C-X-C motif chemokine 10-like

Continued Table 4.5

c68387_g1_i1	1.73E-12	2.66E-09	0.00	0.00	0.00	0.00	0.00	0.00	0.21	10.26	VHSV-induced protein
c49628_g1_i1	5.56E-11	6.51E-08	0.00	0.14	0.06	0.00	0.00	0.06	0.65	4.84	E3 ubiquitin-protein ligase rnf213-like
c2104_g1_i1	1.59E-09	1.19E-06	0.00	0.00	0.00	0.00	0.00	0.06	0.05	6.15	interleukin-17f-like
c5208_g2_i1	1.74E-09	1.25E-06	0.00	0.00	0.00	0.00	0.00	0.00	0.53	29.63	interferon alpha-inducible protein 27-like protein 2-like
c50760_g1_i1	4.33E-09	2.66E-06	0.00	0.00	0.00	0.00	0.00	0.00	0.64	6.46	E3 ubiquitin-protein ligase trim21-like

Note, mm: mock group, mv: NNV challenged group; FDR: false discovery rate; * Fold-change was ratio of mv48h to mm48h, - represents not available because value of mm48h was 0.

Of these genes elicited by NNV, the pro-inflammatory cytokine of interleukin-17f-like (*il17fl*), tumor necrosis factor alpha (*tnfα*) and its rapidly induced tumor necrosis factor alpha-induced protein 3 (*tnfaip3*), were strongly elicited after virus infection. For *il17fl*, there was no expression at all the four time points in the mock samples, and even at 6 hpi in virus challenged samples. Then, its expression began to increase at 12 and 24 hpi, but remained at very low levels of 0.06 and 0.05, respectively. However, at 48 hpi, the expression level was increased to 6.15. This indicates that *il17fl* is a late response gene. Il17fl belongs to the Il17 cytokine family, which consists of Il17a, Il17b, Il17c, Il17d, Il17e and Il17f. They bind to their corresponding receptors to activate downstream pathways to produce a variety of molecules. These include cytokines (Tnfα), chemokines and anti-microbial peptides. The Il17 cytokines are derived from a variety of cell types including epithelial cells. Of these, epithelial cells seem to be the cell sources of Il17c and Il17e, but possibly not Il17f (Gu *et al.* 2013). From this study, the *il17fl* was also expressed in epithelial cells and the possible reason is unknown. It is worth noting that Il17f and Il17fl are closely related. Meanwhile, in the tongue sole, *il17fl* has also been found to be expressed in the brain, heart, gill, muscle, spleen, intestine, kidney, liver and blood under normal conditions. However, its expression is significantly increased in the spleen and kidney only after viral infection but not bacterial infection (Chi & Sun 2015). It is still unknown if *il17fl* is expressed in epithelial cells of tongue sole because the authors probably did not examine the expression in the skin.

It is worthy to note that one of the products induced by Il17 is the chemokine which not only promotes migration of leukocyte under both normal and inflammatory conditions, but also regulate the differentiation of recruited cells and acts as a bridge between innate and adaptive immune responses. The chemokine family comprises four subfamilies according to the patterns of two cysteines at the N terminal: C, CC, CXC and CX3C in

mammals (Alejo & Tafalla 2011). In this study, several cytokine genes, induced by NNV infection, belonged to the chemokine family, including CC chemokine, C-X-C motif chemokine 6-like (*cxcl6*) and C-X-C motif chemokine 10-like (*cxcl10*), which generally act as chemoattractant for the migration of several types of cells. Of these, CC chemokines were totally undetectable in the mock. It was induced in NNV challenged samples but with very low expression: 0.15, 0.23 and 3.21 at 6, 12 and 24 hpi, respectively, while its expression was dramatically increased to 36.69 at 48 hpi. The *cxcl6* and *cxcl10* have similar expression patterns to CC chemokine, remaining very low level in the mock and even in the virus challenged samples before 24 hpi, and considerably increasing at 48 hpi. There are dozens of CC chemokine genes. In channel and blue catfish, it seems the majority of CC chemokines are expressed in many organs and tissues including skin which contains epithelial cells, and some are induced after bacterial infection. Study of expression of *cxcl10* in catfish have shown that it is constitutively expressed in many tissues and seem to lead to a conclusion that it could have no function against bacterial infection (Alejo & Tafalla 2011). However, this study shows that it is induced by NNV in epithelial cells of Asian seabass, which indicates that it may have a role in immune response against viral infection.

Expectedly, *tnfα* and its induced *tnfap3*, their expressions were detectable at all the mock samples but remained at very low levels, and were only significantly increased after 12 hpi in the virus challenged samples. A similar result was also obtained in the epithelial cells of fathead minnow challenged with ranavirus (Holopainen *et al.* 2012), and in the half-smooth tongue sole challenged with *V. anguillarum* (Zhang *et al.* 2015). Tnfα, a type II transmembrane glycoprotein, could be the first secreted cytokine to bind its receptors (p55 and p75) to activate nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), as well as induce apoptosis. Strangely, the expression of caspase-8 (*casp8*), the

key coordinator of TNFα induced apoptosis, did not change in this study. This is in contrast with a previous study in which the CP of NNV could up-regulate expression of *casp8* (Guo *et al.* 2003). The possible reasons could deserve further investigation. The activated NF-κB translates to initiate the production of IFNs and inflammatory cytokines (Workenhe *et al.* 2010). At the same time, the NF-κB signalling pathway and TNF-mediated apoptosis are tightly controlled by Tnfaip3, which was rapidly induced by Tnfα. Dysfunction of Tnfaip3 leads to susceptibility to several human pathogens (Ando *et al.* 2013; Giordano *et al.* 2014). In the current study, *tnfα* and *tnfaip3* were detected with very similar patterns of low expressions in the mock and up-regulation after NNV infection. Future study of manipulation of both genes could increase the resistance to NNV in Asian seabass.

In addition, several genes involved in the interferon signaling pathway were identified. Stimulator of interferon gene (*sting*) was only significantly induced after 24 hpi and with a fold-change of 29.39 at 48 hpi. Sting is localized in the endoplasmic reticulum (ER) and is a key player in innate immunity. It is identified as an adaptor acting downstream of the mitochondria antiviral signaling (MAVS) protein and upstream of transcription factors such as Irf3 to induce type I interferon by activating TANK-binding kinase 1 (*tbk1*) (Ivashkiv & Donlin 2014). Interestingly, Sting was also induced in grass carp by grass carp reovirus (GCRV) and over-expression of Sting in grass carp cells could inhibit the yield of GCRV (Feng *et al.* 2014). Whether this gene has a similar function of inhibiting NNV in Asian seabass needs further investigation. Sting activates Tbk1, which phosphorylates interferon regulator factor 3 (*ifr3*), a key player regulating many facets of innate and adaptive immune response. Expectedly, *irf3* was sensitively induced as early as 6 hpi and with a fold-change of 137.11 at 48 hpi. Notably, *irf3* was also induced in the orange-spotted grouper by NNV and experimentally demonstrated to inhibit the replication of NNV but not Singapore grouper iridovirus (SGIV), which is a DNA virus (Huang *et al.*

2015). This could indicate that over-expression of Irf3 in Asian seabass should suppress NNV replication. The activated Ifr3 forms homodimers, which translocate to the nucleus and subsequently recruit other transcriptional factors to initiate INF transcription and some Isgs (Workenhe *et al.* 2010). In this study, *ifn-α1* was significantly induced at 48 hpi with a fold-change of 63.29. After the production of IFN, the IFN molecules bind to receptors in the cell surface to activate signal cascade mediated by the Janus kinase signal transducer and activator of transcription (JAK-STAT) complex, resulting in the transcription of hundreds of interferon stimulated genes (Isgs) (Schneider *et al.* 2014).

In this experiment, several Isgs, including interferon-induced very large GTPase 1-like (*gvin1*), interferon alpha-inducible protein 27-like protein 2-like (*ifi27l2*), interferon stimulated protein 15 (*isg15*) and other Isgs, were strongly induced with fold changes from 33.35 to 1916.6 at 48 hpi. *Isg15* is induced and becomes covalently conjugated to a variety of cellular proteins upon IFN elicitation. ISG15 contains ubiquitin-like domains and plays essential roles in the ISGylation process for target proteins like Jak, Stat and Mx, by conjugated with three ubiquitin-related enzymes E1, E2 and E3, which are also induced by IFN (Workenhe *et al.* 2010). In addition, Isg15, E3 ubiquitin-protein ligase trim39-like and probably E3 ubiquitin-protein ligase herc4-like were up-regulated as early as 6 hpi, but E3 ubiquitin-protein ligase trim21-like and E3 ubiquitin-protein ligase RNF213-like were only significantly up-regulated after 24 hpi of NNV infection in epithelial cells of Asian seabass. Over-expression of zebrafish ISG15 in an epithelioma papulosum cyprini (EPC) cell line conferred resistance to both RNA and DNA viruses, and induced viperin (Langevin *et al.* 2013). Interestingly, immunoprecipitation suggested that ISG50 of Atlantic salmon could bind to the nucleoprotein of infectious salmon anemia virus (ISAV) (Rokenes *et al.* 2007), which could provide clues for the antiviral mechanism of ISG15.

Not surprisingly, Mx, one of the best-known antiviral proteins stimulated by type I IFN in many vertebrates, had low levels of expression in the mock. In the virus challenged samples, its expression levels were initially low even at 12 hpi , but were significantly increased at 24 and 48 hpi with fold changes of 8.83 and 73.88, respectively. This is consistent with the fact that it is a downstream gene in the IFN signaling pathway. Mx is a dynamin like protein with relatively large molecular weight and contains a highly conserved GTPase domain at its N-terminal, which is responsible for the intrinsic GTPase activity. Although its antiviral mechanism is still unknown, Mx could block intracellular transport of viruses by binding viral nucleocapsids (Workenhe *et al.* 2010). Mx seems to interact with the coat protein of NNV to form a complex of CP-Mx in orange-spotted grouper, resulting the attenuation of viral replication (Lin *et al.* 2006).

Furthermore, the expression levels of two heat shock protein (Hsp) family members, *hsp30* and *hsp70*, were also increased by 1022.23 and 473.67 times at 48 hpi, respectively, after NNV infection. Hsps are critical to maintaining cellular homeostasis and protecting cells from various stresses by facilitating nascent protein folding or keeping the solubility of denatured proteins (Hartl 1996). Nevertheless, the relation between Hsps and viruses is complicated. For example, Hsp70 could influence immune response to suppress some virus infections, but could be exploited by other viruses to support their replication (Kim & Oglesbee 2012). The exact role of Hsp70 in the interaction between Asian seabass and NNV is still unknown and further studies are necessary. There is limited information about Hsp30 in virus infection and functional analysis will be essential in the future.

Lastly, several patterns of gene expression (Table 4.5) were also observed. For most of the genes involved in innate immunity, their expressions in mock remained very low level, but increased to high levels after NNV infection. For example, the expression levels of 23 out of 30 genes were less than one in the mock, but substantially increased

after virus infection. In particular, the expressions of seven genes, including CC chemokine, *il17fl* and *ifi27l2*, were totally undetected in the mock. Of these genes, Il17fl and Ifi27l2 were undetectable even at 6 hpi in the NNV challenged samples. This indicates that maintaining these genes in activation state might not be required at normal condition, whereas activation and maintaining them at a certain level could be necessary after detection of the presence of a virus. Furthermore, most gene expressions remained at low levels at the early stage of infection at 6 and 12 hpi, respectively. For example, expression levels of 21 genes were less than one, and three of them (VSHV-induced protein, Ifi27l2 and E3 ubiquitin-protein ligase trim21-like) were not detectable. However, at the late stage of infection at 24 and 48 hpi, their expression levels were increased considerably. Several genes including, *tnfα*, *irf3*, *rtp3*, *hsp30*, *hsp70* and *isg15*, were significantly induced as early as 6 hpi, which demonstrates that these genes are very sensitive upon virus infection. This suggests that they are the early response genes and could play important roles in the immune response. Furthermore, several genes, like *mx*, *viperin* and *gvin1*, were only significantly up-regulated after 24 hpi, which indicates that they are late response genes. This also implies that these genes were not sensitively induced and may not mount enough immune reactions against virus infection before multiple rounds of virus replications and massive cellular damage in the infected cells had already occurred. It is possible to manipulate some key late response genes so that they are sensitively increased to a certain level of expression before or even shortly after virus infection. This could result in suppression of virus replication or even eliminating the virus from infected cells. This has been demonstrated in studies of Mx (Lin *et al.* 2006) and Viperin (Wang *et al.* 2014a) in fish against several viruses. In addition, the observation of strong induction of innate immunity is in line with previous studies in diverse fish species in response to various pathogens, indicating that they probably share a similar defense mechanism and innate immunity is conserved across fish species.

4.3.5. Molecular characterization of *viperin*

4.3.5.1. *Viperin* is highly induced in vitro and in vivo

For the DEGs involved in innate immunity, a very well-known ISG was *viperin* (virus inhibitory protein, endoplasmic reticulum-associated, interferon-inducible). The gene sequence was 2724 bp long, consisting of six exons and five introns (Fig. 4.4). It contained an OFR of 1056 bp, encoding a protein of 351 aa. Further sequence analysis revealed two SSRs in *viperin*, which could be used as markers for association study. One was (CT)4 in the second intron, the other one was (GA)4aa(GA)5 in the fifth intron (Fig. 4.4).

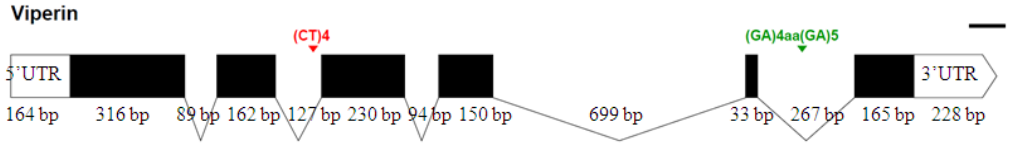


Fig. 4.4 Schematic representation of gene structure of *viperin* of Asian seabass.Each black bar represents an exon. Each number represents the length of a UTR, exon or intron. Two SSRs are indicated by the red and green marks.

RNA-seq data revealed that its expression levels remained almost undetectable in the mock and at 6 and 12 hpi in the virus challenged samples. However, its expression levels increased from 0.04 and 0 to 1.93 and 56.81 at 24 and 48 hpi, respectively (Table 4.5). It had a fold change of 763.86 at 48 hpi, was one of genes with the highest fold changes. Further qPCR analysis confirmed that it was also strongly induced in almost all the examined organs and tissues of NNV-infected Asian seabass at 5 dpc, including brain, eye, fin, heart, intestine, kidney, liver, muscle and skin, except in spleen (Fig. 4.5).

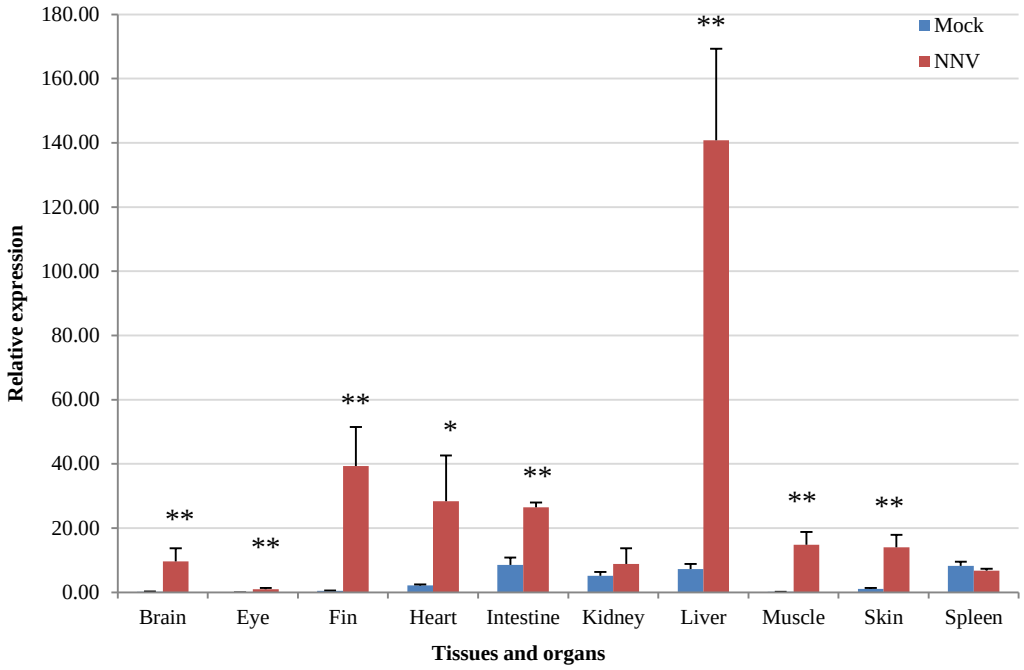


Fig. 4.5 Expression pattern of *viperin* from mock and NNV-challenged samples at 5 days-post challenge of Asian seabass. Statistical significance: *, $p < 0.05$; **, $p < 0.01$.

The multifunctional Viperin seems to have the ability to limit a large variety of both DNA and RNA viruses, through blocking the release of virus particles (Helbig & Beard 2014). However, several pieces of evidence demonstrated that some viruses could overcome Viperin-mediated inhibition and utilize Viperin to promote virus replication instead (Helbig & Beard 2014). In addition, Wang and his team demonstrated that the *viperin* of crucian carp (*Carassius carassius*) could suppress the replication of GCRV (Wang *et al.* 2014a). High induction of Asian seabass *viperin* in vitro and in vivo indicates that it could play a role in Asian seabass-NNV interaction. Whether Asian seabass Viperin inhibits or promotes NNV multiplication remains to be further investigated.

4.3.5.2. Cellular localization of fusion protein Viperin-GFP

Cellular localization of a protein could be an important hint indicating its potential function. To determine the cellular localization of Viperin, construct of Viperin-pEGFP-N1 was built and expressed in Asian seabass epithelial cells. GFP signals, detected by

Confocal, demonstrated that the fusion protein of Viperin-EGFP was localized in the cytosol (Fig. 4.6). This is roughly consistent with a previous study, in which the Viperin of crucian carp was localized in ER (Wang *et al.* 2014b). In this study, ER marker was not used. Therefore, it was impossible to determine the subcellular localization. Since Viperin is a highly conserved protein, Asian seabass Viperin should also be located in the ER. Furthermore, NNV virions were located in the smooth ER of the infected cells (Crane & Hyatt 2011), the possibility of direct interaction between NNV and Viperin exists.

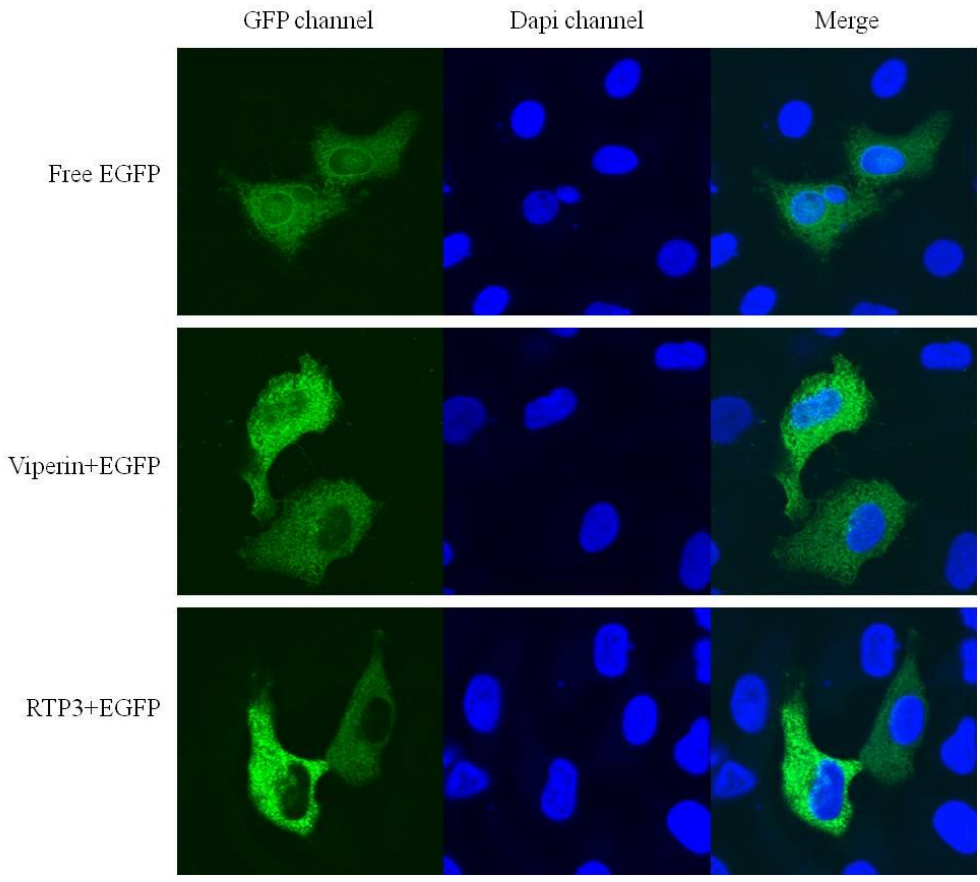


Fig. 4.6 Cellular localization of fusion protein of Viperin-EGFP and RTP3-EGFP in Asian seabass epithelial cells.

4.3.5.3. Association of *viperin* with VNN disease in multiple families of Asian seabass

Viperin was highly induced in vivo and in vitro after NNV infection and had the possibility of direct interaction with NNV. This has provided some implications for its

roles in Asian seabass-NNV interaction. Association of Viperin with VNN disease could further support that its roles in Asian seabass-NNV interaction. To assess its association with VNN disease resistance, the SSR of (CT)4 in the second intron was used to genotype an association population, consisting of 651 mortalities and 476 surviving fish from 43 families (as described in Chapter 3). After removing missing genotype, 650 mortalities and 470 surviving fish were remained for further analysis. Chi-square test showed association between genotypes (CT)4 and VNN disease was not statistically significant ($p = 0.81$). This could imply that SSR (CT)4 possibly has no influence on the expression of *viperin*, which in turn could affect the resistance trait. Therefore, the SSR of (CT)4 could be under natural selection by disease. All in all, this does not necessarily demonstrate that Viperin is not related to VNN disease. It is also possible that the SSR (GA)4aa(GA)5 in the fifth intron is associated with VNN disease. This deserves further investigation. In addition, SNPs in exons could impact on trait expression (Sachidanandam *et al.* 2001). Therefore, future study could focus on the identification of SNPs in the exons of *viperin* and its association with VNN disease resistance in Asian seabass.

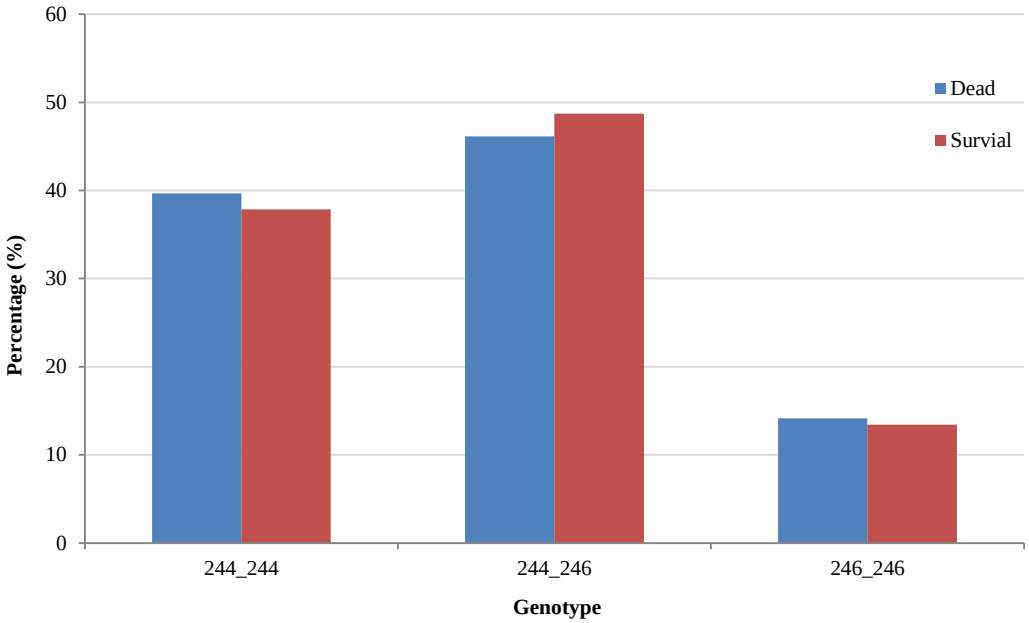


Fig. 4.7 No association between the SSR of (CT)4 in the second intron of *viperin* and VNN disease resistance in the association mapping population of Asian seabass.

4.3.6. Molecular characterization of *RTP3*

4.3.6.1. *RTP3* is highly induced in vitro and in vivo in Asian seabass

Surprisingly, a virus-responsive gene (VSG), *rtp3*, was extraordinarily induced by the NNV. The gene sequence was 2117 bp long, consisting of two exons and one intron (Fig. 4.8). The ORF was 543 bp long, encoding 180 aa. Further sequence analysis showed that there were two SSRs in this gene. One SSR was (AAAAG)4 in the intron, the other was (GT)19tt(GT)4 in the 3'UTR (Fig. 4.8), which could be used for association study. The expression of *rtp3* was less than 0.4 and remained nearly unchanged in the mock. However, the expression levels were significantly increased as early as 6 hpi after NNV infection, and the fold changes were 3.18, 27.95, 79.65 and 758.79 at 6, 12, 24 and 48 hpi, respectively (Table 4.5). It remained as one of the most highly expressed genes after NNV infection. Further gene expression analysis by qPCR showed that *rtp3* was significantly up-regulated in almost all the examined organs and tissues of NNV-infected Asian seabass at 5 dpc, including brain, eye, fin, heart, intestine, kidney, liver, muscle and skin, but was

suppressed in spleen (Fig. 4.9) Meanwhile, *rtp3* was also observed to be the most strongly induced gene in several fish species, including Atlantic salmon and Atlantic cod, during infection of NNV, SAV and piscine reovirus (PRV) (Krasnov *et al.* 2011; Krasnov *et al.* 2013; Johansen *et al.* 2015).

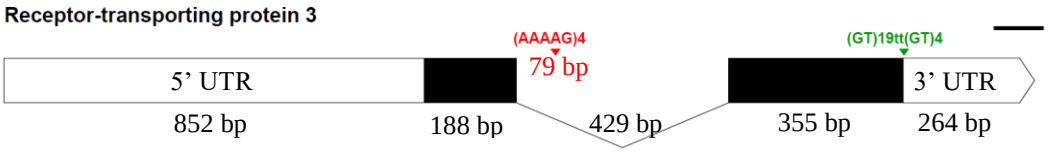


Fig. 4.8 Schematic representation gene structure of *rtp3* in Asian seabass. Each black bar represents an exon. Each number represents the length of a UTR, exon or intron; Two SSRs are indicated by the red and green marks.

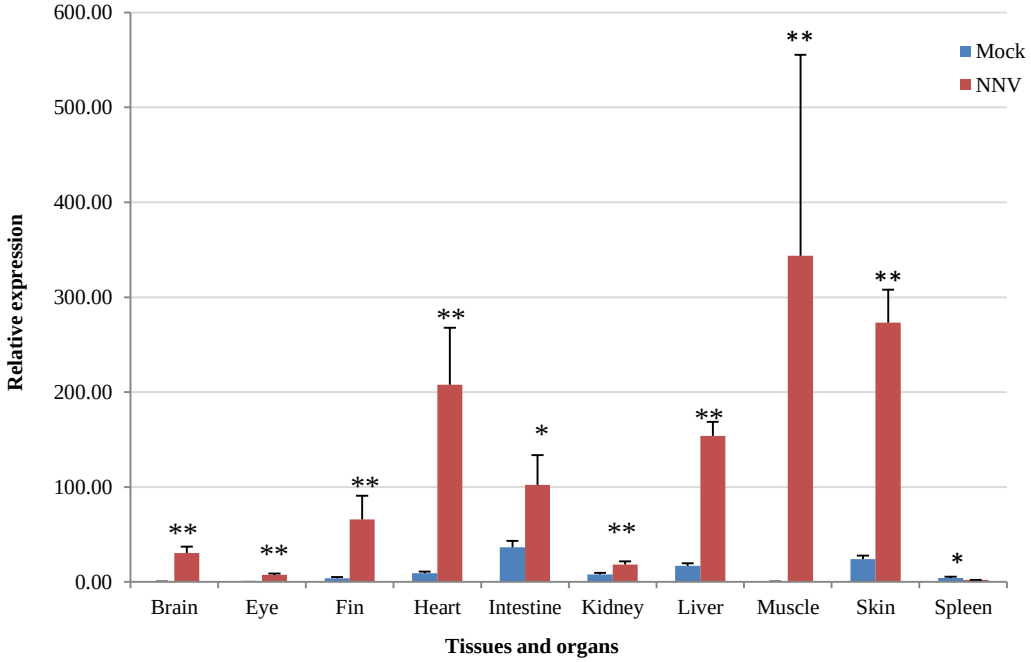


Fig. 4.9 Expression pattern of *rtp3* from mock and NNV-challenged samples at 5 days-post challenge in Asian seabass. Statistical significance: *, $p < 0.05$; **, $p < 0.01$.

4.3.6.2. Cellular localization of fusion protein RTP3-GFP

Identification of the cellular localization of a protein could provide some indications for its potential function. The fusion protein of RTP3-EGFP was expressed in Asian seabass epithelial cells. Detection of GFP signals showed that Rtp3 was localized in

the cytosol (Fig. 4.6). The exact subcellular localization of Rtp3 was not known due to the lack of subcellular markers. Further studies could focus on the identification of its subcellular localization. However, the interaction between Rtp3 and NNV could exist due to the fact that NNV virions were located in the smooth ER of the infected cells (Crane & Hyatt 2011).

4.3.6.3. Association of *rtp3* with VNN disease in multiple families of Asian seabass

The significant induction of Rtp3 in vivo and in vitro aroused my interest in whether Rtp3 is associated with VNN disease. Association study revealed that both SSRs were significantly associated with VNN disease. As shown in Fig. 4.10, the genotypes of SSR (AAAAG)4 in the intron was significantly correlated with VNN disease ($p = 0.0002$). Further analysis of the second SSR, (GT)19tt(GT)4 in the 3'UTR, demonstrated that its genotypes were also significantly associated with VNN disease in Asian seabass ($p = 0$). As shown in Fig. 4.11, the genotype of 410_417 could increase the survival rate, and this could be used in MAS in Asian seabass disease resistance breeding program. It is well-known that SSRs in 3'UTRs may regulate gene expression by influencing gene transcription, mRNA stability and degradation, translocation and translation, eventually leading to phenotype changes (Li *et al.* 2004). Understanding how this SSR influences *rtp3* requires its functional analysis in a future study. In addition, SNPs in exons could impact on trait expression (Sachidanandam *et al.* 2001). Therefore, future study could focus on the identification of SNPs in the exons of *rtp3* and its association with VNN disease resistance in Asian seabass. Furthermore, to my surprise, Rtp family members are accessory proteins responsible for trafficking odorant receptor proteins to the cell-surface membrane in mammals (Behrens *et al.* 2006). Unexpectedly, *rtp3* was demonstrated as IFN-α inducible and suppressing the cancer cell growth of human hepatocellular carcinoma (Zhou *et al.* 2007). These results strongly indicate that *rtp3* might be a novel resistance gene which may

play a very important role in the interaction between host and NNV, and its exact roles in the interaction between fish and virus require further investigation.

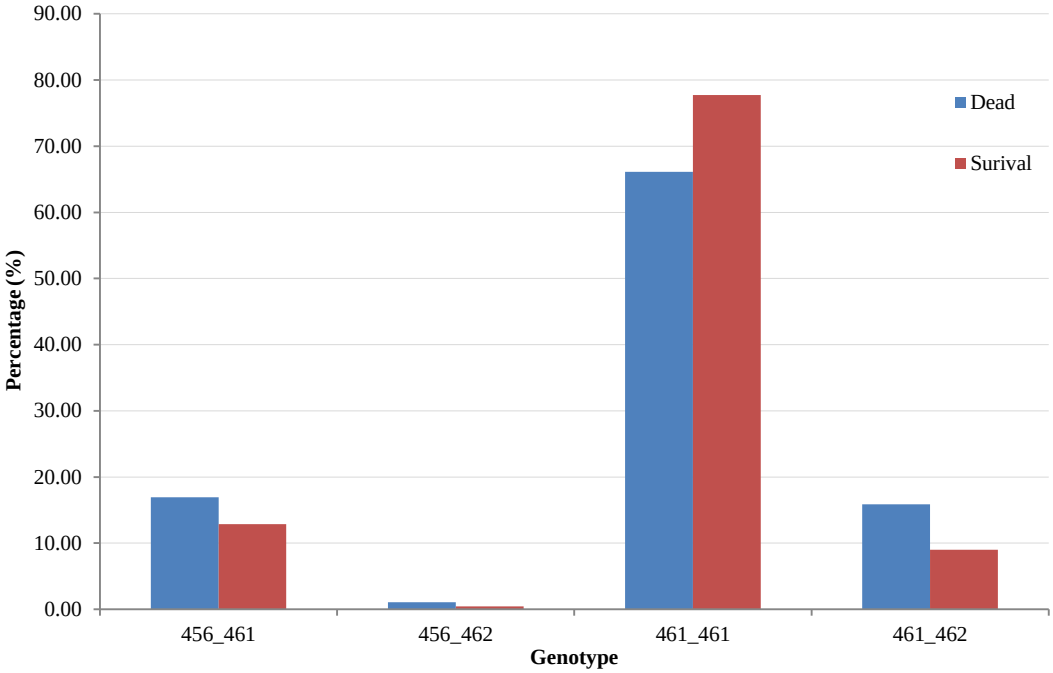


Fig. 4.10 Association between the SSR of (AAAAG)4 in the intron of *rtp3* and VNN disease resistance in the association population of Asian seabass.

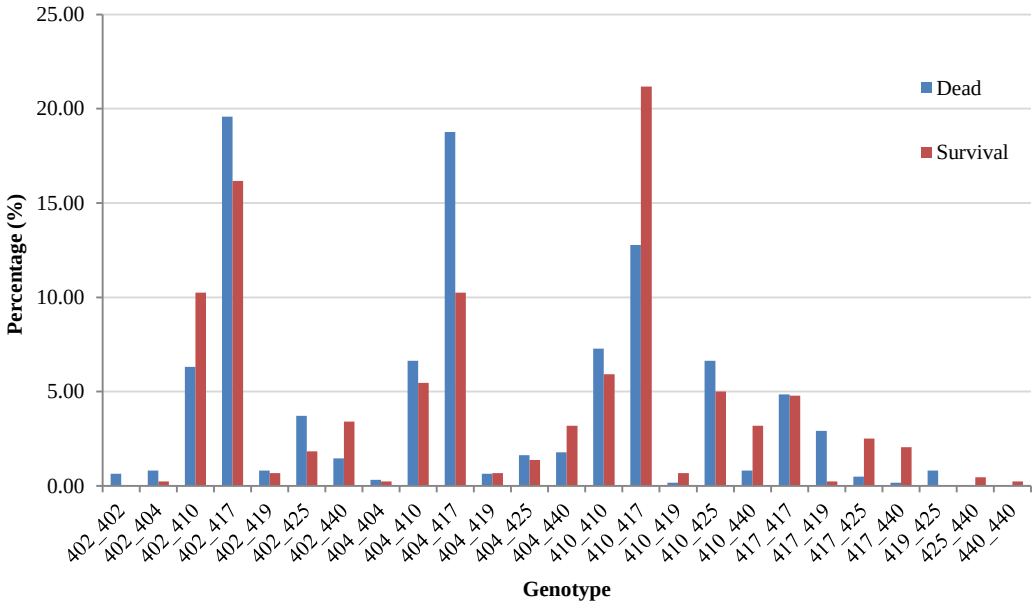


Fig. 4.11 Association between the SSR of (GT)19tt(GT)4 in the 3'UTR of *rtp3* and VNN disease resistance in the association population of Asian seabass.

4.3.7. qRT-PCR verification for the DEGs

To validate the expression profiling of DEGs identified by RNA-seq, fifteen up-regulated and four down-regulated genes (Appendix Table A3) were selected, and primer pairs were designed based on the corresponding assembled contigs. Single PCR bands produced by the primer pairs were detected (data not shown), indicating that the *de novo* assembled transcriptome was reliable. Thereafter, relative expression levels of the selected genes could be determined by qPCR. Comparing the fold changes of the selected genes determined by RNA-seq and qPCR showed that the two data sets are largely consistent. For example, the fold changes in RNA-seq and qPCR, respectively, of several up-regulated genes were: 9118.63 and 12675.93 for protein asteroid homolog 1-like, 758.79 and 679.12 for *rtp3*, 473.67 and 309.61 for *hsp70*, 137.11 and 118.85 for *irf3*, 73.88 and 59.77 for *mx* (Table 4.6). However, there were relatively big difference between RNA-seq and qPCR data sets for a few other genes, such as 1916.6 and 154.95 for *gvin1*, 1022.23 and 108.38 for *hsp30*, 295.35 and 85.4 for *cxcl6*, respectively. In addition, the fold changes of the down-regulated genes in RNA-seq and qPCR data sets, respectively, at 48 hpi, were 12.07 and 9.55 for integrin beta-3-like, 11.29 and 3.14 for macrophage-capping protein, 18.98 and 2.61 for growth-regulated alpha isoform x1, 8.76 and 6.08 for myocilin-like (Table 4.6). In general, the expression levels of selected up-regulated and down-regulated genes determined by qPCR were roughly consistent with the data obtained from RNA-seq, with some small variations, indicating that expression profiling of DEGs determined by RNA-seq is reliable and accurate.

Table 4.6 Fold changes of selected genes determined by qPCR for validation of RNA-seq of Asian seabass.

Gene id	RNA-seq				qPCR				Change	Sequence description	Sequence length (bp)
	6hpi	12hpi	24hpi	48hpi	6hpi	12hpi	24hpi	48hpi			
c244_g1_i1	0.78	1.07	3.34	12.07	1.1	0.2	5.56	9.55	Down-regulated	integrin beta-3-like	3353
c8093_g1_i1	0.62	0.88	2.07	11.29	1.57	0.73	0.91	3.14	Down-regulated	macrophage-capping protein	1466
c51925_g1_i1	0.55	0.75	3.02	18.98	1.54	1.32	1.75	2.61	Down-regulated	growth-regulated alpha isoform x1	896
c9997_g1_i1	0.89	1.39	2.32	8.76	1.12	0.24	3.25	6.08	Down-regulated	myocilin-like	1936
c35231_g1_i1	0.53	9.21	110.43	9118.63	0.6	9.35	238.05	12675.93	Up-regulated	protein asteroid homolog 1-like	2360
c3737_g1_i1	1.35	4.21	1.329*	1916.6	9.43	7.72	5.35	154.95	Up-regulated	interferon-induced very large gtpase 1-like	6102
c11620_g1_i1	4.01	6.38	170.15	1022.23	0.78	9.07	8.62	108.38	Up-regulated	heat shock protein 30-like	830
c11709_g1_i1	3.18	27.95	79.65	758.79	0.44	8.34	37.8	679.12	Up-regulated	receptor-transporting protein 3-like	1422
c12544_g2_i1	3.7	3.77	32.06	622.63	2.14	14.59	2.95	259.18	Up-regulated	urokinase plasminogen activator surface receptor-like	1088
c12370_g3_i1	1.41	8.14	163.82	473.67	1.18	19.65	64.5	309.61	Up-regulated	heat shock protein 70	1092
c144_g1_i1	4.63	9.37	77.31	295.35	0.55	5.45	11.21	85.4	Up-regulated	C-X-C motif chemokine 6-like	822
c12542_g9_i1	4.02	56.96	28.09	137.11	1.98	117.57	23.38	118.85	Up-regulated	interferon regulatory factor 3	2797
c65346_g1_i1	1.18	1.97	4.88	131.13	0.95	2.16	2.82	34.59	Up-regulated	interferon-induced protein with tetratricopeptide repeats 5-like	2142
c8372_g1_i1	1.08	1.34	8.83	73.88	2.15	2.41	5.34	59.57	Up-regulated	interferon inducible mx protein	2912
c7898_g1_i1	0*	1.05	50.33	56.81*	1.95	5.78	3.63	236.86	Up-regulated	viperin	1302
c38898_g1_i1	0.15*	0.23*	3.21*	36.68*	2.35	9.47	22.21	317.58	Up-regulated	CC chemokine	842
c5512_g2_i1	3.01	3.4	3.72	33.35	0.92	1.66	4.05	32.88	Up-regulated	interferon stimulated protein 15	1349
c5208_g2_i1	0*	0*	0.53*	29.62*	0.88	0.81	1.14	50.76	Up-regulated	interferon alpha-inducible protein 27-like protein 2-like	359
c16863_g1_i1	0.92	2.11	2.52	29.39	2.19	1.57	2.81	15.69	Up-regulated	stimulator of interferon genes	1356

Note: * expression data from NNV challenged samples, fold change was not calculated because the expression levels in the mock were 0

4.3.8. Identification of microsatellites in Asian seabass transcriptome

SSRs are highly polymorphic, co-dominant and widely distributed in the whole genome. They are extensively used in evolutionary studies and QTL mapping in MAS programs (Chistiakov *et al.* 2006). SSRs in protein-coding genes and their UTRs may regulate gene expression by transcription and translation, change gene products, and eventually lead to phenotype changes (Li *et al.* 2004). Hence, this type of SSR with putative function should be exposed to much stronger selection pressures compared with those located in the intergenic regions (Li *et al.* 2004). To carry out SSR profiling in the transcriptome of Asian seabass, the 65536 longest transcripts were submitted to the software GMATo (Wang *et al.* 2013). A total of 24816 SSRs (Table 4.7) were found in 17932 (27.36%) transcripts with motifs from di- to hexa-nucleotides and repeats of more than five times. It is worth to note that most of 17932 transcripts harbored more than one SSR and/or multiple types of SSRs in different positions. The most abundant type was di-nucleotide, numbering 19707 (79.44%). This was more than four times of the number of tri-nucleotides (4559, 18.38%). The third most abundant type was tera-nucleotide (508, 2.05%), and pentra- and hexa-nucleotides numbered 24 and 9, respectively (Table 4.7). This data set reveals that more than one quarter of the longest transcripts harbour SSRs, and could provide valuable resources for SSR mining and facilitating QTL mapping in Asian seabass.

Table 4.7 SSRs identified from transcriptome of Asian seabass.

Motif type	Repeat number									Total	%
	5	6	7	8	9	10	11	12	>12		
di-	8132	3282	2098	2007	2424	1495	261	8	0	19707	79.44
tri-	2326	1408	795	27	1	2	0	0	0	4559	18.38
tera-	440	60	3	1	1	1	1	0	1	508	2.05
penta-	13	3	4	1	1	0	0	1	1	24	0.10
hexa-	3	2	1	1	1	0	1	0	0	9	0.04
total	10914	4755	2901	2037	2428	1498	263	9	2	24807	100
%	44.00	19.17	11.69	8.21	9.79	6.04	1.06	0.04	0.01	100	

4.4. Conclusion

In conclusion, the Asian seabass transcriptome, consisting of 89026 transcripts, was *de novo* assembled. This could provide information for gene analysis, novel gene identification and even functional genome study for Asian seabass. A total of 251 DEGs in response to NNV infection in Asian seabass were identified. A number of various genes relevant to innate immunity were identified, including *rtp3*, *viperin*, *ifr3*, *ifn*, *hsp30* and *70*, which were strongly induced by NNV. This could set foundation for understanding the molecular interaction between Asian seabass and NNV. The cytosol localization, strong induction in vivo and in vitro after NNV infection and association with VNN disease resistance of *rtp3*, could provide some hints that *rtp3* is a disease related gene. These results present an overview of the response of Asian seabass to NNV infection from the transcriptome level, promoting our comprehensive understanding of the complexity underlying the host's defense against virus infection. SSRs in the transcriptome also provide valuable resources for QTL mapping, which could improve disease resistance, eventually benefiting the whole Asian seabass industry.

Chapter 5. Concluding remarks and future work

5.1.1. Conclusion

VNN disease causes more than 90% mortality at the larval stage of Asian seabass, and continues to threaten it at the juvenile and adult stages, resulting in huge economic losses in the Asian seabass industry (Shetty *et al.* 2012). The most efficient way to combat diseases is to use disease resistant lines. However, such a line of Asian seabass is not available. In addition, disease resistance is a complex/quantitative trait with polygenetic inheritance. Complex traits are generally considered to be controlled by many genes with small effects, and influenced by environment and gene-environment interaction (Gjedrem & Baranski 2010). Identification of chromosome regions controlling quantitative traits involves development of a mapping population, genome-wide genetic markers, construction of a linkage map and QTL mapping. In this thesis, primary QTL mapping and fine mapping for resistance to VNN disease in Asians seabass were performed, as well as a candidate gene underlying a QTL with the largest effect was identified.

To perform primary QTL mapping, Asian seabass fingerlings of a single back-cross family at 37 dph were challenged with NNV at a concentration of 9×10^6 $TCID_{50}$/ml for two hours. A panel of 330 mortalities and 190 surviving fingerlings was genotyped and a linkage map consisting of 145 markers covering 24 linkage groups was constructed. QTL analysis was conducted using interval mapping. Thirteen and 10 QTL were identified for VNN resistance and survival time, respectively. One significant QTL, spanning 3 cM in LG20, was identified for both VNN resistance and survival time, with PVE of 2.2-4.1% for resistance and 2.2-3.3% for survival time, respectively. These results reveal that VNN disease resistance in Asian seabass is polygenetic and controlled by many genes with small effects. This is in consistent with the classical assumption that disease resistance is a

complex and polygenetic trait. Furthermore, this study could also provide basic information for fine mapping.

The confidence interval of QTL identified from primary mapping is relatively large, which limits identification of possible genes underlying the QTL. Fine mapping QTL could narrow down the relatively large QTL region by increasing the number of genetic markers in the linkage map. To conduct fine mapping and further identify corresponding genes, 6425 SNPs from 85 dead and 94 surviving individuals were generated using GBS. A high-density linkage map consisting of 24 LGs containing 3000 (2852 SNPs and 148 SSRs) markers, with an average interval distance of 1.28 cM, was constructed. One significant and three suggestive QTL in three LGs with PVE of 8.3 to 11.0% were identified for resistance. Two significant and two suggestive QTL in four LGs with PVE of 7.8 to 10.9% were detected for survival time. Further dissection of the QTL with the highest PVE leaded to identification of *pcdhac2* as a possible candidate gene. Association study on multiple families demonstrated that a six bp indel in the second intron of *pcdhac2* was significantly associated with disease resistance. In addition, qPCR analysis showed that its expression was significantly up-regulated in several organs and tissues of Asian seabass after NNV infection. These results could further support *pcdhac2* being a candidate gene controlling disease resistance in Asian seabass. The QTL identified in this study could help us to understand the genetic architecture of VNN disease resistance in Asian seabass. Furthermore, identification of the candidate gene *pcdhac2* could improve the understanding of possible disease resistance mechanisms. Moreover, the markers tightly linked to QTL could have the potential to be implemented in MAS in disease resistance program in Asian seabass.

Besides identification of QTL which could promise to increase VNN disease resistance in Asian seabass, transcriptome profiling analysis after virus infection is

essential for an overall understanding of the host-virus interaction. This in turn could provide critical information for identification of disease resistance genes. In addition, identification of SSRs in genes could provide valuable resources for QTL mapping and association studies. To investigate the transcriptome profiling after NNV infection, Asian seabass epithelial cells were challenged with NNV. Libraries of eight mRNA samples (6, 12, 24, 48 hours post-inoculation) of mock and NNV-challenged cells, were constructed and sequenced using RNA-seq. The transcriptome of Asian seabass was *de novo* assembled, consisted of 89026 transcripts with a N50 of 2617 bp. Further analysis identified 251 DEGs in response to NNV infection. Of these genes, 30 were involved in innate immunity. The top up-regulated genes included *rtp3*, *viperin*, *irf3* and other genes. Two genes, *rtp3* and *viperin*, were selected for further molecular characterization. qPCR revealed that both genes are strongly induced in many organs and tissues after NNV infection. Cellular localization analysis demonstrated that both proteins were localized in the cytosol, which could provide some hints for their possible function. Furthermore, one SSR in the 3'UTR of *rtp3* was significantly associated with VNN disease resistance in an association study of multiple families of Asian seabass. Molecular characterization and association study of *rtp3* could indicate it is a possible novel disease resistance gene and its SSRs could be applied in MAS program as a selection marker. The data suggest that abundant and diverse genes correspond to NNV infection. This could strength our understanding of Asian seabass-VNN interaction. The genetic markers identified in transcriptome could benefit QTL mapping and association study in Asian seabass.

5.1.2. Future work

The QTL identified in this study have small to moderate effect. This could be due to the single family used in this study, which could not have the major QTL. Moreover,

these QTL have not been validated in the other family, and could be family specific QTL. Therefore, 1) validation of these QTL in other families and perform QTL mapping in multiple families are necessary. In addition, the QTL mapped in this study could only explain a small fraction of total phenotypic variation. More importantly, the LD between marker and QTL will decay at various rate, depending on the recombination rate of marker and QTL and the generations passed. This will compromise the power of MAS, which is based on LD between marker and QTL. As a consequence, 2) GS could provide solution to these problems. Functional analysis of disease related genes are essential to understand the mechanism of disease resistance, hence, 3) identification of gene functions of *rtp3*, *viperin* and *pcdhac2* will be carried out.

Bibliography

Alejo A. & Tafalla C. (2011) Chemokines in teleost fish species. *Developmental & Comparative Immunology* **35**, 1215-22.

Aljanabi S.M. & Martinez I. (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* **25**, 4692-3.

Ando M., Sato Y., Takata K., Nomoto J., Nakamura S., Ohshima K., Takeuchi T., Orita Y., Kobayashi Y. & Yoshino T. (2013) A20 (TNFAIP3) deletion in epstein-barr virus-associated lymphoproliferative disorders/lymphomas. *PLoS ONE* **8**, e56741.

Andrews K.R., Good J.M., Miller M.R., Luikart G. & Hohenlohe P.A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics* **17**, 81-92.

Angeles Esteban M. (2012) An overview of the immunological defenses in fish skin. *International Scholarly Research Notices: Immunology* **2012**, 1-29.

Ayllon F., Kjærner-Semb E., Furmanek T., Wennevik V., Solberg M.F., Dahle G., Taranger G.L., Glover K.A., Almén M.S. & Rubin C.J. (2015) The *vgll3* locus controls age at maturity in wild and domesticated Atlantic salmon (*Salmo salar* L.) males. *PLoS Genetics* **11**, e1005628.

Azad I., Jithendran K., Shekhar M., Thirunavukkarasu A. & de la Pena L. (2006) Immunolocalisation of nervous necrosis virus indicates vertical transmission in hatchery produced Asian sea bass (*Lates calcarifer* Bloch)—A case study. *Aquaculture* **255**, 39-47.

Azad I., Shekhar M., Thirunavukkarasu A., Poornima M., Kailasam M., Rajan J., Ali S., Abraham M. & Ravichandran P. (2005) Nodavirus infection causes mortalities in hatchery produced larvae of *Lates calcarifer*: first report from India. *Diseases of Aquatic Organisms* **63**, 113-8.

Baerwald M., Petersen J., Hedrick R., Schisler G. & May B. (2011) A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (*Oncorhynchus mykiss*). *Heredity* **106**, 920-6.

Barlow C. (1997) *The new rural industries: a handbook for farmers and investors. Ed: Hyde, K. Rural Industries Research and Development Corporation, Australia. Available at: http://www.rirdc.gov.au/pub/handbook/barramundi.html*. Rural Industries Research & Development Corp, Barton.

Barson N.J., Aykanat T., Hindar K., Baranski M., Bolstad G.H., Fiske P., Jacq C., Jensen A.J., Johnston S.E. & Karlsson S. (2015) Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528**, 405-8.

Behrens M., Bartelt J., Reichling C., Winnig M., Kuhn C. & Meyerhof W. (2006) Members of RTP and REEP gene families influence functional bitter taste receptor expression. *Journal of Biological Chemistry* **281**, 20650-9.

Bellance R. & de Saint-Aurin D.G. (1988) L'encéphalite virale du loup de mer. *Caraibes Medical* **2**, 105-44.

Berger A. (2010) Molecular analysis of the *Oryzias latipes* (Medaka) transcriptome. p. 165. Free University of Berlin, Berlin.

Bloch B., Gravningen K. & Larsen J. (1991) Encephalomyelitis among turbot associated with a picornavirus-like agent. *Diseases of Aquatic Organisms* **10**, 65-70.

Boopathi N.M. (2012) *Genetic mapping and marker assisted selection: basics, practice and benefits*. Springer, New York.

Brown L. (1993) *Aquaculture for veterinarians. Fish husbandry and medicine*. Butterworth-Heinemann Limited, Oxford.

Carlson B.M., Onusko S.W. & Gross J.B. (2015) A high-density linkage map for *Astyanax mexicanus* using genotyping-by-sequencing technology. *G3 Genes Genomics Genetics* **5**, 241-51.

Castric J., Thiery R., Jeffroy J., De Kinkelin P. & Raymond J. (2001) Sea bream *Sparus aurata*, an asymptomatic contagious fish host for nodavirus. *Diseases of Aquatic Organisms* **47**, 33-8.

Catchen J.M., Amores A., Hohenlohe P., Cresko W. & Postlethwait J.H. (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3 Genes Genomics Genetics* **1**, 171-82.

Chang S., Ngoh G., Kueh L., Qin Q., Chen C., Lam T. & Sin Y. (2001) Development of a tropical marine fish cell line from Asian seabass (*Lates calcarifer*) for virus isolation. *Aquaculture* **192**, 133-45.

Chen L.J., Su Y.C. & Hong J.R. (2009) Betanodavirus non-structural protein B1: A novel anti-necrotic death factor that modulates cell death in early replication cycle in fish cells. *Virology* **385**, 444-54.

Chenoweth S.F., Hughes J.M., Keenan C.P. & Lavery S. (1998) When oceans meet: a teleost shows secondary intergradation at an Indian–Pacific interface. *Proceedings of the Royal Society of London B: Biological Sciences* **265**, 415-20.

Chi H. & Sun L. (2015) Comparative study of four interleukin 17 cytokines of tongue sole *Cynoglossus semilaevis*: Genomic structure, expression pattern, and promoter activity. *Fish & Shellfish Immunology* **47**, 321-30.

Chistiakov D.A., Hellemans B. & Volckaert F.A. (2006) Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture* **255**, 1-29.

Chong S., Ngoh G. & Chew-Lim M. (1990) Study of three tissue culture viral isolates from marine foodfish. *Singapore Journal of Primary Industries* **18**, 54-7.

Chong S., Ngoh G., Ng M. & Chu K. (1987) Growth of lymphocystis virus in a sea bass, *Lates calcarifer* (Bloch 1709) cell line. *Singapore Veterinary Journal* **11**, 78-85.

Churchill G.A. & Doerge R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-71.

Collard B., Jahufer M., Brouwer J. & Pang E. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* **142**, 169-96.

Conesa A., Gotz S., Garcia-Gomez J.M., Terol J., Talon M. & Robles M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-6.

Crane M. & Hyatt A. (2011) Viruses of fish: an overview of significant pathogens. *Viruses* **3**, 2025-46.

Cui Y., Wang H., Qiu X., Liu H. & Yang R. (2015) Bayesian analysis for genetic architectures of body weights and morphological traits using distorted markers in Japanese flounder *Paralichthys olivaceus*. *Marine Biotechnology* **17**, 693-702.

Davey J.W., Hohenlohe P.A., Etter P.D., Boone J.Q., Catchen J.M. & Blaxter M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

Dekkers J.C. & Hospital F. (2002) The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* **3**, 22-32.

Dhert P., Lavens P., Duray M. & Sorgeloos P. (1990) Improved larval survival at metamorphosis of Asian seabass (*Lates calcarifer*) using ω3-HUFA-enriched live food. *Aquaculture* **90**, 63-74.

Dhert P., Lavens P. & Sorgeloos P. (1992) State of the art of Asian seabass *Lates calcarifer* larviculture. *Journal of the World Aquaculture Society* **23**, 317-29.

Diouf J. (2009) How to feed the world in 2050. FAO's Director-General's Statements.

Doerge R.W. & Churchill G.A. (1996) Permutation tests for multiple loci affecting a quantitative character. *Genetics* **142**, 285-94.

Dorson M., Quillet E., Hollebecq M., Torhy C. & Chevassus B. (1995) Selection of rainbow trout resistant to viral haemorrhagic septicaemia virus and transmission of resistance by gynogenesis. *Veterinary Research* **26**, 361-8.

Duarte C.M., Marbá N. & Holmer M. (2007) Rapid domestication of marine species. *Science* **316**, 382-3.

Ellis A. (2001) Innate host defense mechanisms of fish against viruses and bacteria. *Developmental & Comparative Immunology* **25**, 827-39.

Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S. & Mitchell S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379.

FAO (2014) Global Aquaculture Producti on Volume and Value Statistics Database. Update to 2012. Available at http://www.fao.org/3/a-i3720e.pdf. Access at 20 April, 2016.

FAO (2016) Food and Agriculture Organization of the United Nations Statistics (FAO). Cultured aquatic species information programme - *Lates calcarifer* (Block, 1790). Available at http://www.fao.org/fishery/culturedspecies/Lates_calcarifer/en Access at 20 April, 2016.

Feng X., Yang C., Zhang Y., Peng L., Chen X., Rao Y., Gu T. & Su J. (2014) Identification, characterization and immunological response analysis of stimulator of interferon gene (STING) from grass carp *Ctenopharyngodon idella*. *Developmental & Comparative Immunology* **45**, 163-76.

Fenner B., Du Q., Goh W., Thiagarajan R., Chua H. & Kwang J. (2006a) Detection of betanodavirus in juvenile barramundi, *Lates calcarife*r (Bloch), by antigen capture ELISA. *Journal of Fish Diseases* **29**, 423-32.

Fenner B.J., Goh W. & Kwang J. (2006b) Sequestration and protection of double-stranded RNA by the betanodavirus B2 protein. *Journal of Virology* **80**, 6822-33.

Fenner B.J., Goh W. & Kwang J. (2007) Dissection of double-stranded RNA binding protein B2 from betanodavirus. *Journal of Virology* **81**, 5449-59.

Fenner B.J., Thiagarajan R., Chua H.K. & Kwang J. (2006c) Betanodavirus B2 is an RNA interference antagonist that facilitates intracellular viral RNA accumulation. *Journal of Virology* **80**, 85-94.

Fjalestad K.T., Gjedrem T. & Gjerde B. (1993) Genetic improvement of disease resistance in fish: an overview. *Aquaculture* **111**, 65-74.

Fuji K., Hasegawa O., Honda K., Kumasaka K., Sakamoto T. & Okamoto N. (2007) Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* **272**, 291-5.

Fuji K., Kobayashi K., Hasegawa O., Coimbra M.R.M., Sakamoto T. & Okamoto N. (2006) Identification of a single major genetic locus controlling the resistance to lymphocystis disease in Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* **254**, 203-10.

Geldermann H. (1975) Investigations on inheritance of quantitative characters in animals by gene markers I. Methods. *Theoretical and Applied Genetics* **46**, 319-30.

Geng X., Sha J., Liu S., Bao L., Zhang J., Wang R., Yao J., Li C., Feng J., Sun F., Sun L., Jiang C., Zhang Y., Chen A., Dunham R., Zhi D. & Liu Z. (2015) A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. *BMC Genomics* **16**, 196-207.

Georges M. (2007) Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. *Annual Review Genomics and Human Genetics* **8**, 131-62.

Gilbey J., Verspoor E., Mo T.A., Jones C. & Noble L. (2006) Identification of genetic markers associated with Gyrodactylus salaris resistance in Atlantic salmon *Salmo salar*. *Diseases of Aquatic Organisms* **71**, 119.

Giordano M., Roncagalli R., Bourdely P., Chasson L., Buferne M., Yamasaki S., Beyaert R., van Loo G., Auphan-Anezin N., Schmitt-Verhulst A.M. & Verdeil G. (2014) The tumor necrosis factor alpha-induced protein 3 (TNFAIP3, A20) imposes a brake on antitumor activity of CD8 T cells. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 11115-20.

Gjedrem T. & Baranski M. (2010) *Selective breeding in aquaculture: an introduction*. Springer Netherlands, Dordrecht, Netherlands.

Gjedrem T. & Robinson N. (2014) Advances by selective breeding for aquatic species: a review. *Agricultural Sciences* **5**, 1152-8.

Gjedrem T., Robinson N. & Rye M. (2012) The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture* **350**, 117-29.

Glazebrook J. & Campbell R. (1987) Diseases of barramundi (*Lates calcarifer*) in Australia: a review. In: *Management of Wild and Cultured seabass/barramundi*, pp. 204-6, Darwin, NT, Australia.

Glazebrook J., Heasman M. & Beer S. (1990) Picorna-like viral particles associated with mass mortalities in larval barramundi, *Lates calcarifer* Bloch. *Journal of Fish Diseases* **13**, 245-9.

Glazer A.M., Killingbeck E.E., Mitros T., Rokhsar D.S. & Miller C.T. (2015) Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. *G3 Genes Genomics Genetics* **5**, 1463-72.

Goddard M.E. & Hayes B.J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**, 381-91.

Gonen S., Baranski M., Thorland I., Norris A., Grove H., Arnesen P., Bakke H., Lien S., Bishop S. & Houston R. (2015) Mapping and validation of a major QTL affecting resistance to pancreas disease (*Salmonid alphavirus*) in Atlantic salmon (*Salmo salar*). *Heredity* **115**, 405-18.

Gonen S., Lowe N.R., Cezard T., Gharbi K., Bishop S.C. & Houston R.D. (2014) Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genomics* **15**, 166-82.

Gorjanc G., Cleveland M.A., Houston R.D. & Hickey J.M. (2015) Potential of genotyping-by-sequencing for genomic selection in livestock populations. *Genetic Selection Evolution* **47**, 1-14.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R. & Zeng Q. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-52.

Gu C., Wu L. & Li X. (2013) IL-17 family: cytokines, receptors and signaling. *Cytokine* **64**, 477-85.

Guo Y.X., Wei T., Dallmann K. & Kwang J. (2003) Induction of caspase-dependent apoptosis by betanodaviruses GGNNV and demonstration of protein α as an apoptosis inducer. *Virology* **308**, 74-82.

Gutierrez A.P., Lubieniecki K.P., Fukui S., Withler R.E., Swift B. & Davidson W.S. (2014) Detection of quantitative trait loci (QTL) related to grilsing and late sexual maturation in Atlantic salmon (*Salmo salar*). *Marine Biotechnology* **16**, 103-10.

Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., MacManes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., LeDuc R.D., Friedman N. & Regev A. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**, 1494-512.

Haddad-Boubaker S., Bigarré L., Bouzgarou N., Megdich A., Baud M., Cabon J. & Chéhida N.B. (2013) Molecular epidemiology of betanodaviruses isolated from sea bass and sea bream cultured along the Tunisian coasts. *Virus Genes* **46**, 412-22.

Harker N., Rampling L.R., Shariflou M.R., Hayden M.J., Holton T.A., Morell M.K., Sharp P.J., Henry R.J. & Edwards K.J. (2001) Microsatellites as markers for Australian wheat improvement. *Australian Journal of Agricultural Research* **52**, 1121-30.

Hartl F.U. (1996) Molecular chaperones in cellular protein folding. *Nature* **381**, 571-9.

Hayes B. & Goddard M. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-29.

Hegedűs Z., Zakrzewska A., Ágoston V.C., Ordas A., Rácz P., Mink M., Spaink H.P. & Meijer A.H. (2009) Deep sequencing of the zebrafish transcriptome response to mycobacterium infection. *Molecular Immunology* **46**, 2918-30.

Helbig K.J. & Beard M.R. (2014) The role of viperin in the innate antiviral response. *Journal of Molecular Biology* **426**, 1210-9.

Henning F., Lee H.J., Franchini P. & Meyer A. (2014) Genetic mapping of horizontal stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using RAD markers for dense linkage mapping. *Molecular Ecology* **23**, 5224-40.

Hick P., Schipp G., Bosmans J., Humphrey J. & Whittington R. (2011) Recurrent outbreaks of viral nervous necrosis in intensively cultured barramundi (*Lates calcarifer*) due to horizontal transmission of betanodavirus and recommendations for disease control. *Aquaculture* **319**, 41-52.

Hodneland K., Garcia R., Balbuena J., Zarza C. & Fouz B. (2011) Real-time RT-PCR detection of betanodavirus in naturally and experimentally infected fish from Spain. *Journal of Fish Diseases* **34**, 189-202.

Hohenlohe P.A., Day M.D., Amish S.J., Miller M.R., Kamps-Hughes N., Boyer M.C., Muhlfeld C.C., Allendorf F.W., Johnson E.A. & Luikart G. (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology* **22**, 3002-13.

Holopainen R., Tapiovaara H. & Honkanen J. (2012) Expression analysis of immune response genes in fish epithelial cells following ranavirus infection. *Fish & Shellfish Immunology* **32**, 1095-105.

Houston R.D., Davey J.W., Bishop S.C., Lowe N.R., Mota-Velasco J.C., Hamilton A., Guy D.R., Tinch A.E., Thomson M.L. & Blaxter M.L. (2012) Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics* **13**, 244-68.

Houston R.D., Gheyas A., Hamilton A., Guy D.R., Tinch A.E., Taggart J.B., McAndrew B.J., Haley C.S. & Bishop S.C. (2008) Detection and confirmation of a major QTL affecting resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon (*Salmo salar*). *Developmental Biology* **132**, 199-204.

Houston R.D., Haley C.S., Hamilton A., Guy D.R., Mota-Velasco J.C., Gheyas A.A., Tinch A.E., Taggart J., Bron J. & Starkey W. (2010) The susceptibility of Atlantic salmon fry to freshwater infectious pancreatic necrosis is largely explained by a major QTL. *Heredity* **105**, 318-27.

Houston R.D., Taggart J.B., Cezard T., Bekaert M., Lowe N.R., Downing A., Talbot R., Bishop S.C., Archibald A.L., Bron J.E., Penman D.J., Davassi A., Brew F., Tinch

A.E., Gharbi K. & Hamilton A. (2014) Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* **15**, 90-112.

Huang B., Tan C., Chang S., Munday B., Mathew J., Ngoh G. & Kwang J. (2001) Detection of nodavirus in barramundi, Lates calcarifer (Bloch), using recombinant coat protein-based ELISA and RT-PCR. *Journal of Fish Diseases* **24**, 135-42.

Huang Y., Huang X., Cai J., OuYang Z., Wei S., Wei J. & Qin Q. (2015) Identification of orange-spotted grouper (*Epinephelus coioides*) interferon regulatory factor 3 involved in antiviral immune response against fish RNA virus. *Fish & Shellfish Immunology* **42**, 345-52.

Huang Y., Huang X., Yan Y., Cai J., Ouyang Z., Cui H., Wang P. & Qin Q. (2011) Transcriptome analysis of orange-spotted grouper (*Epinephelus coioides*) spleen in response to Singapore grouper iridovirus. *BMC Genomics* **12**, 556-71.

Ivashkiv L.B. & Donlin L.T. (2014) Regulation of type I interferon responses. *Nature Reviews Immunology* **14**, 36-49.

Jerry D.R. (2013) *Biology and culture of Asian aeabass Lates calcarifer*. CRC Press, Boca Raton.

Johansen L.-H., Thim H.L., Jørgensen S.M., Afanasyev S., Strandskog G., Taksdal T., Fremmerlid K., McLoughlin M., Jørgensen J.B. & Krasnov A. (2015) Comparison of transcriptomic responses to pancreas disease (PD) and heart and skeletal muscle inflammation (HSMI) in heart of Atlantic salmon (*Salmo salar* L). *Fish & Shellfish Immunology* **46**, 612-23.

Johnson K.N., Johnson K.L., Dasgupta R., Gratsch T. & Ball L.A. (2001) Comparisons among the larger genome segments of six nodaviruses and their encoded RNA replicases. *Journal of General Virology* **82**, 1855-66.

Kärber G. (1931) 50% end point calculation. *Arch Exp Pathol Pharmakol* **162**, 480-3.

Kearsey M.J.P. & Harpal S. (1996) *The genetical analysis of quantitative traits*. Springer, New York.

Keenan C. & Salini J. (1990) The genetic implications of mixing barramundi stocks in Australia. In: *Introduced and translocated fishes and their ecological effects*, pp. 145-50. Australian Government Publishing Service, Canberra.

Kim M.Y. & Oglesbee M. (2012) Virus-heat shock protein interaction and a novel axis for innate antiviral immunity. *Cells* **1**, 646-66.

Kongchum P., Palti Y., Hallerman E.M., Hulata G. & David L. (2010) SNP discovery and development of genetic markers for mapping innate immune response genes in common carp (*Cyprinus carpio*). *Fish & Shellfish Immunology* **29**, 356-61.

Krasnov A., Kileng O., Skugor S., Jorgensen S.M., Afanasyev S., Timmerhaus G., Sommer A.I. & Jensen I. (2013) Genomic analysis of the host response to nervous necrosis virus in Atlantic cod (*Gadus morhua*) brain. *Molecular Immunology* **54**, 443-52.

Krasnov A., Timmerhaus G., Schiotz B.L., Torgersen J., Afanasyev S., Iliev D., Jorgensen J., Takle H. & Jorgensen S.M. (2011) Genomic survey of early responses to viruses in Atlantic salmon, *Salmo salar* L. *Molecular Immunology* **49**, 163-74.

Laghari M.Y., Lashari P., Zhang X., Xu P., Narejo N.T., Liu Y., Mehboob S., Al-Ghanim K., Zhang Y. & Sun X. (2014) Mapping QTLs for swimming ability related traits in *Cyprinus carpio* L. *Marine Biotechnology* **16**, 629-37.

Lallias D., Gomez-Raya L., Haley C., Arzul I., Heurtebise S., Beaumont A., Boudry P. & Lapegue S. (2009) Combining two-stage testing and interval mapping strategies to detect QTL for resistance to bonamiosis in the European flat oyster *Ostrea edulis*. *Marine Biotechnology* **11**, 570-84.

Lander E.S. & Botstein D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-99.

Langevin C., van der Aa L.M., Houel A., Torhy C., Briolat V., Lunazzi A., Harmache A., Bremont M., Levraud J.P. & Boudinot P. (2013) Zebrafish ISG15 exerts a strong antiviral activity against RNA and DNA viruses and regulates the interferon response. *Journal of Virology* **87**, 10025-36.

Li C., Waldbieser G., Bosworth B., Beck B.H., Thongda W. & Peatman E. (2014) SNP discovery in wild and domesticated populations of blue catfish, *Ictalurus furcatus*, using genotyping-by-sequencing and subsequent SNP validation. *Molecular Ecology Resources* **14**, 1261-70.

Li C., Zhang Y., Wang R., Lu J., Nandi S., Mohanty S., Terhune J., Liu Z. & Peatman E. (2012) RNA-seq analysis of mucosal immune responses reveals signatures of intestinal barrier disruption and pathogen entry following *Edwardsiella ictaluri* infection in channel catfish, *Ictalurus punctatus*. *Fish & Shellfish Immunology* **32**, 816-27.

Li G., Zhao Y., Liu Z., Gao C., Yan F., Liu B. & Feng J. (2015) *De novo* assembly and characterization of the spleen transcriptome of common carp (*Cyprinus carpio*) using Illumina paired-end sequencing. *Fish & Shellfish Immunology* **44**, 420-9.

Li H., Bradbury P., Ersoz E., Buckler E.S. & Wang J. (2011) Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. *PLoS ONE* **6**, e17573.

Li Y.C., Korol A.B., Fahima T. & Nevo E. (2004) Microsatellites within genes: structure, function, and evolution. *Molecular Biology and Evolution* **21**, 991-1007.

Lin C.H., Christopher John J.A., Lin C.H. & Chang C.Y. (2006) Inhibition of nervous necrosis virus propagation by fish Mx proteins. *Biochemical and Biophysical Research Communications* **351**, 534-9.

Liu B.H. (1997) *Statistical genomics: linkage, mapping, and QTL analysis*. CRC press, Boca Raton.

Liu F., Sun F., Xia J.H., Li J., Fu G.H., Lin G., Tu R.J., Wan Z.Y., Quek D. & Yue G.H. (2014a) A genome scan revealed significant associations of growth traits with a major QTL and GHR2 in tilapia. *Scientific Reports* **4**, 7256.

Liu S., Sun L., Li Y., Sun F., Jiang Y., Zhang Y., Zhang J., Feng J., Kaltenboeck L., Kucuktas H. & Liu Z. (2014b) Development of the catfish 250K SNP array for genome-wide association studies. *BMC Research Notes* **7**, 135-46.

Liu S., Vallejo R.L., Gao G., Palti Y., Weber G.M., Hernandez A. & Rexroad C.E., 3rd (2015) Identification of single-nucleotide polymorphism markers associated with cortisol response to crowding in rainbow trout. *Marine Biotechnology* **17**, 328-37.

Liu Z.J. & Cordes J.F. (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**, 1-37.

Livak K.J. & Schmittgen T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* **25**, 402-8.

López-Muñoz A., Sepulcre M.P., García-Moreno D., Fuentes I., Béjar J., Manchado M., Álvarez M.C., Meseguer J. & Mulero V. (2012) Viral nervous necrosis virus persistently replicates in the central nervous system of asymptomatic gilthead seabream and promotes a transient inflammatory response followed by the infiltration of IgM+ B lymphocytes. *Developmental & Comparative Immunology* **37**, 429-37.

Lu M.W., Ngou F.H., Chao Y.M., Lai Y.S., Chen N.Y., Lee F.Y. & Chiou P.P. (2012) Transcriptome characterization and gene expression of *Epinephelus spp* in endoplasmic reticulum stress-related pathway during betanodavirus infection in vitro. *BMC Genomics* **13**, 651-62.

Lu X., Wang H., Liu B. & Xiang J. (2013) Three EST-SSR markers associated with QTL for the growth of the clam *Meretrix meretrix* revealed by selective genotyping. *Marine Biotechnology* **15**, 16-25.

Mackay T.F., Stone E.A. & Ayroles J.F. (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**, 565-77.

Maeno Y., de la Peña L.D. & Cruz-Lacierda E.R. (2004) Mass mortalities associated with viral nervous necrosis in hatchery-reared sea bass *Lates calcarifer* in the Philippines. *Japan Agricultural Research Quarterly* **38**, 69-73.

Magnadottir B. (2010) Immunological control of fish diseases. *Marine Biotechnology* **12**, 361-79.

Manin B.O. & Ransangan J. (2011) Experimental evidence of horizontal transmission of Betanodavirus in hatchery-produced Asian seabass, *Lates calcarifer* and brown-marbled grouper, *Epinephelus fuscoguttatus* fingerling. *Aquaculture* **321**, 157-65.

Mao M.-G., Jiang J.-L., Perálvarez-Marín A., Wang K.-J. & Lei J.-L. (2013) Characterization of the Mx and hepcidin genes in *Epinephelus akaara* asymptomatic carriers of the nervous necrosis virus. *Aquaculture* **408**, 175-83.

Mathew G. (2009) *Taxonomym, identification and biology of seabass (Lates calcarifer)*. Fisheries and Aquaculture Department, Kochi.

McBeath A., Aamelfot M., Christiansen D.H., Matejusova I., Markussen T., Kaldhusdal M., Dale O.B., Weli S.C. & Falk K. (2015) Immersion challenge with low and highly virulent infectious salmon anaemia virus reveals different pathogenesis in Atlantic salmon, *Salmo salar* L. *Journal of Fish Diseases* **38**, 3-15.

Mikalsen A.B., Torgersen J., Alestrom P., Hellemann A.-L., Koppang E.-O. & Rimstad E. (2004) Protection of Atlantic salmon *Salmo salar* against infectious pancreatic necrosis after DNA vaccination. *Diseases of Aquatic Organisms* **60**, 11-20.

Milbourne D., Isidore E., van Os H., Bryan G., van Eck H., Rousselle-Bourgeois F., Ritter E., Bakker J. & Waugh R. (2000) A 1260 point genetic linkage map of potato chromosome 1: paving the way for ultra high density genetic linkage maps in crop species. *Scottish Crop Research Institute Annual Report*, 75-8.

Miles C. & Wayne M. (2008) Quantitative trait locus (QTL) analysis. *Nature Education* **1**, 208.

Moen T., Baranski M., Sonesson A.K. & Kjøglum S. (2009) Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. *BMC Genomics* **10**, 368-81.

Moen T., Sonesson A.K., Hayes B., Lien S., Munck H. & Meuwissen T.H. (2007) Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*): comparing survival analysis with analysis on affected/resistant data. *BMC Genetics* **8**, 53-65.

Moen T., Torgersen J., Santi N., Davidson W.S., Baranski M., Ødegård J., Kjøglum S., Velle B., Kent M. & Lubieniecki K.P. (2015) Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics* **200**, 1313-26.

Mohan M., Nair S., Bhagwat A., Krishna T., Yano M., Bhatia C. & Sasaki T. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding* **3**, 87-103.

Mooi R.D. & Gill A. (1995) Association of epaxial musculature with dorsal-fin pterygiophores in acanthomorph fishes, and its phylogenetic significance. *Bulletin of the Natural History Museum. Zoology Series* **61**, 121-37.

Mori K.-I., Nakai T., Muroga K., Arimoto M., Mushiake K. & Furusawa I. (1992) Properties of a new virus belonging to Nodaviridae found in larval striped jack (*Pseudocaranx dentex*) with nervous necrosis. *Virology* **187**, 368-71.

Mori K, Nakai T, Muroga K, Mukuchi T & T K. (1991) A viral disease in hatchery-reared larvae and juveniles of redspotted grouper. *Fish Pathology* **26**, 209-10.

Morishita H. & Yagi T. (2007) Protocadherin family: diversity, structure, and function. *Current Opinion in Cell Biololgy* **19**, 584-92.

Mortazavi A., Williams B.A., McCue K., Schaeffer L. & Wold B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621-8.

Munday B., Langdon J., Hyatt A. & Humphrey J. (1992) Mass mortality associated with a viral-induced vacuolating encephalopathy and retinopathy of larval and juvenile barramundi, *Lates calcarifer* Bloch. *Aquaculture* **103**, 197-211.

Mutz K.-O., Heilkenbrinker A., Lönne M., Walter J.-G. & Stahl F. (2013) Transcriptome analysis using next-generation sequencing. *Current Opinion in Biotechnology* **24**, 22-30.

Myles S. (2013) Improving fruit and wine: what does genomics have to offer? *Trends in Genetics* **29**, 190-6.

Nagalakshmi U., Wang Z., Waern K., Shou C., Raha D., Gerstein M. & Snyder M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344-9.

Nakai T., Sugaya T., Nishioka T., Mushiake K. & Yamashita H. (2009) Current knowledge on viral nervous necrosis (VNN) and its causative betanodaviruses. *The Israeli Journal of Aquaculture* **61**, 198-207.

Nelson J.S. (2006) *Fishes of the World*. John Wiley & Sons, New Jersey.

Nguyen N. & Ponzoni R. (2006) Perspectives from Agriculture: Advances in livestock breeding-Implications for aquaculture genetics. *NAGA, WorldFish Center Quarterly* **29**, 39-45.

Ødegård J., Baranski M., Gjerde B. & Gjedrem T. (2011a) Methodology for genetic evaluation of disease resistance in aquaculture species: challenges and future prospects. *Aquaculture Research* **42**, 103-14.

Ødegård J., Gitterle T., Madsen P., Meuwissen T.H., Yazdi M.H., Gjerde B., Pulgarin C. & Rye M. (2011b) Quantitative genetics of taura syndrome resistance in pacific white shrimp (*Penaeus vannamei*): a cure model approach. *Genetics Selection Evolution* **43**, 14-20.

Ødegård J., Sommer A.-I. & Præbel A.K. (2010) Heritability of resistance to viral nervous necrosis in Atlantic cod (*Gadus morhua* L.). *Aquaculture* **300**, 59-64.

Olveira J., Souto S., Dopazo C. & Bandín I. (2013) Isolation of betanodavirus from farmed turbot *Psetta maxima* showing no signs of viral encephalopathy and retinopathy. *Aquaculture* **406**, 125-30.

Otero O. (2004) Anatomy, systematics and phylogeny of both recent and fossil latid fishes (Teleostei, *Perciformes, Latidae*). *Zoological Journal of the Linnean Society* **141**, 81-133.

Ozaki A., Khoo S.-K., Yoshiura Y., Ototake M., Sakamoto T., Dijkstra J.M. & Okamoto N. (2007) Identification of additional quantitative trait loci (QTL) responsible for susceptibility to infectious pancreatic necrosis virus in rainbow trout. *Fish Pathology* **42**, 131-40.

Ozaki A., Sakamoto T., Khoo S., Nakamura K., Coimbra M., Akutsu T. & Okamoto N. (2001) Quantitative trait loci (QTLs) associated with resistance/susceptibility to infectious pancreatic necrosis virus (IPNV) in rainbow trout (*Oncorhynchus mykiss*). *Molecular Genetics and Genomics* **265**, 23-31.

Ozaki A., Yoshida K., Fuji K., Kubota S., Kai W., Aoki J.-y., Kawabata Y., Suzuki J., Akita K. & Koyama T. (2013) Quantitative trait loci (QTL) associated with resistance to a monogenean parasite (*Benedenia seriolae*) in yellowtail (*Seriola quinqueradiata*) through genome wide analysis. *PLoS ONE* **8**, e64987.

Ozsolak F. & Milos P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics* **12**, 87-98.

Panzarin V., Fusaro A., Monne I., Cappellozza E., Patarnello P., Bovo G., Capua I., Holmes E.C. & Cattoli G. (2012) Molecular epidemiology and evolutionary dynamics of betanodavirus in southern Europe. *Infection, Genetics and Evolution* **12**, 63-70.

Parameswaran V., Kumar S.R., Ahmed V.I. & Hameed A.S. (2008) A fish nodavirus associated with mass mortality in hatchery-reared Asian Sea bass, *Lates calcarifer*. *Aquaculture* **275**, 366-9.

Patel R.K. & Jain M. (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* **7**, e30619.

Peterson B.K., Weber J.N., Kay E.H., Fisher H.S. & Hoekstra H.E. (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**, e37135.

Pethiyagoda R. & Gill A.C. (2012) Description of two new species of sea bass (Teleostei: *Latidae: Lates*) from Myanmar and Sri Lanka. *Zootaxa* **3314**, 1-16.

Pierce C.P. (2006) The Next Big Fish. *The Boston Globe Magazine* **22**, 2008.

Poland J.A., Brown P.J., Sorrells M.E. & Jannink J.-L. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* **7**, e32253.

Poland J.A. & Rife T.W. (2012) Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* **5**, 92-102.

Purcell M.K., Marjara I.S., Batts W., Kurath G. & Hansen J.D. (2011) Transcriptome analysis of rainbow trout infected with high and low virulence strains of infectious hematopoietic necrosis virus. *Fish & Shellfish Immunology* **30**, 84-93.

Qian X., Ba Y., Zhuang Q. & Zhong G. (2014) RNA-Seq technology and its application in fish transcriptomics. *Omics: A Journal of Integrative Biology* **18**, 98-110.

Ransangan J. & Manin B.O. (2010) Mass mortality of hatchery-produced larvae of Asian seabass, *Lates calcarifer* (Bloch), associated with viral nervous necrosis in Sabah, Malaysia. *Veterinary Microbiology* **145**, 153-7.

Reyes-Cerpa S., Sandino A.M., Toro-Ascuy D., Reyes-López F., Maisey K. & Imarai M. (2012) *Fish cytokines and immune response*. InTech, Rijeka, Croatia.

Ribeiro F.F. & Qin J.G. (2013) Modelling size-dependent cannibalism in barramundi *Lates calcarifer*: cannibalistic polyphenism and its implication to aquaculture. *PLoS ONE* **8**, e82488.

Risch N. (1992) Genetic linkage: interpreting lod scores. *Science* **255**, 803-4.

Rise M.L., Hall J.R., Rise M., Hori T.S., Browne M.J., Gamperl A.K., Hubert S., Kimball J., Bowman S. & Johnson S.C. (2010) Impact of asymptomatic nodavirus carrier state and intraperitoneal viral mimic injection on brain transcript expression in Atlantic cod (*Gadus morhua*). *Physiological Genomics* **42**, 266-80.

Roberts R.J. (2012) *Fish pathology*. John Wiley & Sons, New Jersey.

Rodriguez-Ramilo S.T., De La Herran R., Ruiz-Rejon C., Hermida M., Fernandez C., Pereiro P., Figueras A., Bouza C., Toro M.A. & Martinez P. (2014) Identification of quantitative trait loci associated with resistance to viral haemorrhagic septicaemia (VHS) in turbot (*Scophthalmus maximus*): a comparison between bacterium, parasite and virus diseases. *Marine Biotechnology* **16**, 265-76.

Rodriguez-Ramilo S.T., Fernandez J., Toro M.A., Bouza C., Hermida M., Fernandez C., Pardo B.G., Cabaleiro S. & Martinez P. (2013) Uncovering QTL for resistance and survival time to *Philasterides dicentrarchi* in turbot (*Scophthalmus maximus*). *Animal Genetics* **44**, 149-57.

Rodriguez-Ramilo S.T., Toro M.A., Bouza C., Hermida M., Pardo B.G., Cabaleiro S., Martinez P. & Fernandez J. (2011) QTL detection for *Aeromonas salmonicida* resistance related traits in turbot (*Scophthalmus maximus*). *BMC Genomics* **12**, 541-50.

Rokenes T.P., Larsen R. & Robertsen B. (2007) Atlantic salmon ISG15: Expression and conjugation to cellular proteins in response to interferon, double-stranded RNA and virus infections. *Molecular Immunology* **44**, 950-9.

Sachidanandam R., Weissman D., Schmidt S.C., Kakol J.M., Stein L.D., Marth G., Sherry S., Mullikin J.C., Mortimore B.J. & Willey D.L. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-33.

Salini J. & Shaklee J. (1988) Genetic structure of barramundi (*Lates calcarifer*) stocks from northern Australia. *Marine and Freshwater Research* **39**, 317-29.

Sánchez C.C., Smith T.P., Wiedmann R.T., Vallejo R.L., Salem M., Yao J. & Rexroad C.E. (2009) Single nucleotide polymorphism discovery in rainbow trout by deep sequencing of a reduced representation library. *BMC Genomics* **10**, 1-8.

Sauvage C., Boudry P., De Koning D.J., Haley C.S., Heurtebise S. & Lapègue S. (2010) QTL for resistance to summer mortality and OsHV-1 load in the Pacific oyster (*Crassostrea gigas*). *Animal Genetics* **41**, 390-9.

Schipp G., Bosmans J. & Humphrey J. (2007) *Barramundi farming handbook*, Darwin.

Schleimer R.P., Kato A., Kern R., Kuperman D. & Avila P.C. (2007) Epithelium: at the interface of innate and adaptive immune responses. *Journal of Allergy and Clinical Immunology* **120**, 1279-84.

Schlote W. (1979) *Breeding and improvement of farm animals:(7th edn)*. McGraw-Hill, New York.

Schneider W.M., Chevillotte M.D. & Rice C.M. (2014) Interferon-stimulated genes: a complex web of host defenses. *Annual Review of Immunology* **32**, 513-45.

Shaykhiev R. & Bals R. (2007) Interactions between epithelial cells and leukocytes in immunity and tissue homeostasis. *Journal of Leukocyte Biology* **82**, 1-15.

Shetty M., Maiti B., Santhosh K.S., Venugopal M.N. & Karunasagar I. (2012) Betanodavirus of marine and freshwater fish: distribution, genomic organization, diagnosis and control measures. *Indian Journal of Virology* **23**, 114-23.

Smith S., Bernatchez L. & Beheregaray L.B. (2013) RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics* **14**, 375-86.

Sneller C., Mather D. & Crepieux S. (2009) Analytical approaches and population types for finding and utilizing QTL in complex plant populations. *Crop Science* **49**, 363-80.

Sonesson A.K. & Meuwissen T. (2009) Testing strategies for genomic selection in aquaculture breeding programs. *Genetic Selection Evolution* **41**, 37.

Srichanun M., Tantikitti C., Utarabhand P. & Kortner T.M. (2013) Gene expression and activity of digestive enzymes during the larval development of Asian seabass (*Lates calcarifer*). *Comparative Biochemistry and Physiology* **165**, 1-9.

Stear M., Nikbakht G., Matthews L. & Jonsson N. (2013) *Breeding for disease resistance in livestock and fish*, Oxfordshire.

Tan C., Huang B., Chang S.F., Ngoh G.H., Munday B., Chen S.C. & Kwang J. (2001) Determination of the complete nucleotide sequences of RNA1 and RNA2 from greasy grouper (*Epinephelus tauvina*) nervous necrosis virus, Singapore strain. *Journal of General Virology* **82**, 647-53.

Tanksley S.D. (1993) Mapping polygenes. *Annual Review of Genetics* **27**, 205-33.

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S. & McCouch S. (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* **11**, 1441-52.

Thevasagayam N.M., Sridatta P.S., Jiang J., Tong A., Saju J.M., Kathiresan P., Kwan H.Y., Ngoh S.Y., Liew W.C. & Kuznetsova I.S. (2015) Transcriptome survey of a marine food fish: Asian seabass (*Lates calcarifer*). *Journal of Marine Science and Engineering* **3**, 382-400.

Thodesen J. & Gjedrem T. (2006) Breeding programs on Atlantic salmon in Norway: lessons learned. In: *WorldFish Center Conference Proceedings,73*, p. 22. WorldFish Center, Penang, Malaysia.

Tian M., Li Y., Jing J., Mu C., Du H., Dou J., Mao J., Li X., Jiao W., Wang Y., Hu X., Wang S., Wang R. & Bao Z. (2015) Construction of a high-density genetic map and quantitative trait locus mapping in the sea cucumber *Apostichopus japonicus*. *Scientific Reports* **5**, 14852.

Tookwinas S. (1989) Larviculture of seabass (*Lates calcarifer*) and grouper (*Epinephelus malabaricus*) in Thailand. In: *Advances in Tropical Aquaculture*, p. 15, French Polynesia.

Tran N.T., Gao Z.X., Zhao H.H., Yi S.K., Chen B.X., Zhao Y.H., Lin L., Liu X.Q. & Wang W.M. (2015) Transcriptome analysis and microsatellite discovery in the blunt snout bream (*Megalobrama amblycephala*) after challenge with *Aeromonas hydrophila*. *Fish & Shellfish Immunology* **45**, 72-82.

Uribe C., Folch H., Enriquez R. & Moran G. (2011) Innate and adaptive immunity in teleost fish: a review. *Veterinary Medicine* **56**, 486-503.

Vallejo R.L., Palti Y., Liu S., Evenhuis J.P., Gao G., Rexroad C.E., 3rd & Wiens G.D. (2014) Detection of QTL in rainbow trout affecting survival when challenged with *Flavobacterium psychrophilum*. *Marine Biotechnology* **16**, 349-60.

Van Ooijen J. (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genetics Research* **93**, 343-9.

Van Ooijen J. & Kyazma B. (2009) MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. *Kyazma BV: Wageningen, Netherlands*.

Van Os H., Andrzejewski S., Bakker E., Barrena I., Bryan G.J., Caromel B., Ghareeb B., Isidore E., de Jong W. & Van Koert P. (2006) Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. *Genetics* **173**, 1075-87.

Van Roy F. & Berx G. (2008) The cell-cell adhesion molecule E-cadherin. *Cell Molecular Life Science* **65**, 3756-88.

Vasta G.R., Nita-Lazar M., Giomarelli B., Ahmed H., Du S., Cammarata M., Parrinello N., Bianchet M.A. & Amzel L.M. (2011) Structural and functional diversity of the lectin repertoire in teleost fish: relevance to innate and adaptive immunity. *Developmental & Comparative Immunology* **35**, 1388-99.

Veerkamp R., Brotherstone S., Engel B. & Meuwissen T. (2001) Analysis of censored survival data using random regression models. *Animal Science* **72**, 1-10.

Vike S., Nylund S. & Nylund A. (2009) ISA virus in Chile: evidence of vertical transmission. *Archives of Virology* **154**, 1-8.

Vinas A., Taboada X., Vale L., Robledo D., Hermida M., Vera M. & Martinez P. (2012) Mapping of DNA sex-specific markers and genes related to sex differentiation in turbot (*Scophthalmus maximus*). *Marine Biotechnology* **14**, 655-63.

Visscher P.M., Thompson R. & Haley C.S. (1996) Confidence intervals in QTL mapping by bootstrapping. *Genetics* **143**, 1013-20.

Voorrips R. (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93**, 77-8.

Wang B., Zhang Y.-B., Liu T.-K., Shi J., Sun F. & Gui J.-F. (2014a) Fish viperin exerts a conserved antiviral function through RLR-triggered IFN signaling pathway. *Developmental & Comparative Immunology* **47**, 140-9.

Wang B., Zhang Y.B., Liu T.K. & Gui J.F. (2014b) Sequence analysis and subcellular localization of crucian carp Carassius auratus viperin. *Fish & Shellfish Immunology* **39**, 168-77.

Wang C.M., Bai Z.Y., He X.P., Lin G., Xia J.H., Sun F., Lo L.C., Feng F., Zhu Z.Y. & Yue G.H. (2011) A high-resolution linkage map for comparative genome analysis and QTL fine mapping in Asian seabass, *Lates calcarifer*. *BMC Genomics* **12**, 174-91.

Wang L., Fan C., Liu Y., Zhang Y., Liu S., Sun D., Deng H., Xu Y., Tian Y., Liao X., Xie M., Li W. & Chen S. (2014c) A genome scan for quantitative trait loci associated with *Vibrio anguillarum* infection resistance in Japanese flounder (*Paralichthys olivaceus*) by bulked segregant analysis. *Marine Biotechnology* **16**, 513-21.

Wang L., Wan Z.Y., Bai B., Huang S.Q., Chua E., Lee M., Pang H.Y., Wen Y.F., Liu P., Liu F., Sun F., Lin G., Ye B.Q. & Yue G.H. (2015a) Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass. *Scientific Reports* **5**, 16358.

Wang L., Xia J.H., Liu X.J., Liu P., Wan Z.Y. & Yue G.H. (2014d) Molecular characterization and mapping of Fgf21 gene in a foodfish species Asian seabass. *PLoS ONE* **9**, e90172.

Wang P., Wang J., Su Y.-Q., Mao Y., Zhang J.-S., Wu C.-W., Ke Q.-Z., Han K.-H., Zheng W.-Q. & Xu N.-d. (2016) Transcriptome analysis of the *Larimichthys crocea* liver in response to *Cryptocaryon irritans*. *Fish & Shellfish Immunology* **48**, 1-11.

Wang W., Hu Y., Ma Y., Xu L., Guan J. & Kong J. (2015b) High-density genetic linkage mapping in turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex-and growth-related regions detection. *PLoS ONE* **10**, e0120410.

Wang X., Lu P. & Luo Z. (2013) GMATo: A novel tool for the identification and analysis of microsatellites in large genomes. *Bioinformation* **9**, 541-4.

Wang Y.D., Rajanbabu V. & Chen J.Y. (2015c) Transcriptome analysis of medaka following epinecidin-1 and TH1-5 treatment of NNV infection. *Fish & Shellfish Immunology* **42**, 121-31.

Wang Z., Gerstein M. & Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57-63.

Ward R., Holmes B. & Yearsley G. (2008) DNA barcoding reveals a likely second species of Asian seabass (barramundi) (*Lates calcarifer*). *Journal of Fish Biology* **72**, 458-63.

Wedemeyer G. (1996) *Physiology of fish in intensive culture systems*. Springer US, New York.

Wiens G.D., Vallejo R.L., Leeds T.D., Palti Y., Hadidi S., Liu S., Evenhuis J.P., Welch T.J. & Rexroad C.E., III (2013) Assessment of genetic correlation between bacterial cold water disease resistance and spleen index in a domesticated population of rainbow trout: Identification of QTL on chromosome omy19. *PLoS ONE* **8**, e75749.

Workenhe S.T., Rise M.L., Kibenge M.J. & Kibenge F.S. (2010) The fight between the teleost fish immune response and aquatic viruses. *Molecular immunology* **47**, 2525-36.

Wu T.D. & Watanabe C.K. (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859-75.

Xia J.H., Lin G., He X., Liu P., Liu F., Sun F., Tu R. & Yue G.H. (2013a) Whole genome scanning and association mapping identified a significant association between growth and a SNP in the IFABP-a gene of the Asian seabass. *BMC Genomics* **14**, 295-305.

Xia J.H., Lin G., He X., Yunping B., Liu P., Liu F., Sun F., Tu R. & Yue G.H. (2014) Mapping quantitative trait loci for omega-3 fatty acids in Asian seabass. *Marine Biotechnology* **16**, 1-9.

Xia J.H., Liu P., Liu F., Lin G., Sun F., Tu R. & Yue G.H. (2013b) Analysis of stress-responsive transcriptome in the intestine of Asian seabass (*Lates calcarifer*) using RNA-Seq. *DNA Research* **20**, 449-60.

Xiang L., He D., Dong W., Zhang Y. & Shao J. (2010) Deep sequencing-based transcriptome profiling analysis of bacteria-challenged *Lateolabrax japonicus* reveals insight into the immune-relevant genes in marine fish. *BMC Genomics* **11**, 1-21.

Xiao S., Wang P., Zhang Y., Fang L., Liu Y., Li J.T. & Wang Z.Y. (2015) Gene map of large yellow croaker (*Larimichthys crocea*) provides insights into teleost genome evolution and conserved regions associated with growth. *Scientific Reports* **5**, 18661.

Xu Y. (2010) *Molecular plant breeding*. CABI, Wallingford, UK.

Yang Y., Yu H., Li H. & Wang A. (2016) Transcriptome profiling of grass carp (*Ctenopharyngodon idellus*) infected with *Aeromonas hydrophila*. *Fish & Shellfish Immunology* **51**, 329-36.

Ye H., Liu Y., Liu X., Wang X. & Wang Z. (2014) Genetic mapping and QTL analysis of growth traits in the large yellow croaker *Larimichthys crocea*. *Marine Biotechnology* **16**, 729-38.

Yoshikoshi K. & Inoue K. (1990) Viral nervous necrosis in hatchery-reared larvae and juveniles of Japanese parrotfish, *Oplegnathus fasciatus* (Temminck & Schlegel). *Journal of Fish Diseases* **13**, 69-77.

You X., Shu L., Li S., Chen J., Luo J., Lu J., Mu Q., Bai J., Xia Q. & Chen Q. (2013) Construction of high-density genetic linkage maps for orange-spotted grouper *Epinephelus coioides* using multiplexed shotgun genotyping. *BMC Genetics* **14**, 113-20.

Young N. (1996) QTL mapping and quantitative disease resistance in plants. *Annual Review of Phytopathology* **34**, 479-501.

Young W.P., Ostberg C.O., Keim P. & Thorgaard G.H. (2001) Genetic characterization of hybridization and introgression between anadromous rainbow trout (*Oncorhynchus mykiss* irideus) and coastal cutthroat trout (*O. clark*i clarki). *Molecular Ecology* **10**, 921-30.

Yu Y., Zhang X., Yuan J., Li F., Chen X., Zhao Y., Huang L., Zheng H. & Xiang J. (2015) Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific White Shrimp *Litopenaeus vannamei*. *Scientific Reports* **5**, 15612.

Yu Z. & Guo X. (2006) Identification and mapping of disease-resistance QTLs in the eastern oyster, *Crassostrea virginica* gmelin. *Aquaculture* **254**, 160-70.

Yue G.H. (2014) Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries* **15**, 376-96.

Yue G.H. & Orban L. (2005) A simple and affordable method for high throughput DNA extraction from animal tissues for polymerase chain reaction. *Electrophoresis* **26**, 3081-3.

Yue G.H., Zhu Z.Y., Lo L.C., Wang C.M., Lin G., Feng F., Pang H.Y., Li J., Gong P. & Liu H.M. (2009) Genetic variation and population structure of Asian seabass (*Lates calcarifer*) in the Asia-Pacific region. *Aquaculture* **293**, 22-8.

Zeng D., Chen X., Xie D., Zhao Y., Yang C., Li Y., Ma N., Peng M., Yang Q., Liao Z., Wang H. & Chen X. (2013) Transcriptome analysis of Pacific white shrimp (*Litopenaeus vannamei)* hepatopancreas in response to taura syndrome virus (TSV) experimental infection. *PLoS ONE* **8**, e57515.

Zhang X., Wang S., Chen S., Chen Y., Liu Y., Shao C., Wang Q., Lu Y., Gong G., Ding S. & Sha Z. (2015) Transcriptome analysis revealed changes of multiple genes involved in immunity in *Cynoglossus semilaevis* during *Vibrio anguillarum* infection. *Fish & Shellfish Immunology* **43**, 209-18.

Zhao C., Fu M., Wang C., Jiao Z. & Qiu L. (2016) RNA-Seq analysis of immune-relevant genes in *Lateolabrax japonicus* during *Vibrio anguillarum* infection. *Fish & Shellfish Immunology* **52**, 57-64.

Zhou X., Popescu N.C., Klein G. & Imreh S. (2007) The interferon-alpha responsive gene TMEM7 suppresses cell proliferation and is downregulated in human hepatocellular carcinoma. *Cancer Genetics Cytogenetics* **177**, 6-15.

Appendices

Table A1 Primer pairs used in Chapter 2

Primer name	Primer sequence (5' - 3')	Target	Product length (bp)	Tm (°C)
RNA1-4F	CGTTACAGCGGGCTCGGAATG	*RNA1*	826	59
RNA1-4R	TTAGCCCAGCCAATGTCGTCAATC			
RNA2-4F	GCTGGGGCATCTGGGACAACTTT	*RNA2*	469	59
RNA2-4R	TCACTGCGCGGAGCTAACGGTAAC			
Lca-tub-F	GGCACTACACAATCGGCAAAGAGA	*α-Tubulin*	144	59
Lca-tub-R	TCAGCAGGGAGGTAAAGCCAGAGC			

Table A2 Primer pairs used in Chapter 3

Primer name	Primer sequence (5' - 3')	Target	Product length (bp)	Tm (°C)
Lca-pcdhαc2-q-F	GCTTATGCCCTGCGAGCCTTTGA	*Pcdhαc2*	166	58
Lca-pcdhαc2-q-R	CATTGCCGGGAGGGTTGACTATCAG			
Lca-pcadhαc2-del-F	CCGTGCCATGCTGTGAGTGC	the six bp indel in *Pcdhαc2*	254	54
Lca-pcadhαc2-del-R	GATGCCGGAGCTGTGTTCTGTTA			
EF1a1-F	GGGCATCTCTACACATATAACCAC	*EF1a*	126	60
EF1a1-R	CAGAGAAACAGGGATGACGA			

Table A3 Primer pairs of qRT-PCR for validation of RNA-seq data in Chapter 4

Gene id	Gene name	Primer name	Primer sequence (5'-3')	Tm (°C)	PCR product length (bp)
c244_g1_i1	integrin beta-3-like	Lca-IB3-q-F	CCACCTCGACCTTCACAAACATCAC	64	99
		Lca-IB3-q-R	TTCGGTGCCAAAGTCCATCATTTCC		
c8093_g1_i1	macrophage-capping protein	Lca-mcp-q-F	CAGCTTCAACAAGGGAGACTGCTTC	64	84
		Lca-mcp-q-R	TCGAAGATGTTGGCTTGAGATCCGA		
c51925_g1_i1	growth-regulated alpha isoform x1	Lca-grai-q-F	CCAACTGAAAGGCAAGAGAGAACGG	64	101
		Lca-grai-q-R	TCTTATTAACTCGTGCCTTCGCCCT		
c9997_g1_i1	myocilin-like	Lca-myol-q-F	CATATGATCCTGGGAATGGGGCGTA	64	82
		Lca-myol-q-R	ACTTCAGGGATAAGGTGGGATGCTG		
c35231_g1_i1	protein asteroid homolog 1-like	Lca-pa-q-F	TGCCCCAGCTCGCCAAGATAGTG	59	177
		Lca-pa-q-R	GCAGCAGCCCAGATGATAGATTGA		
c11709_g1_i1	receptor-transporting protein 3-like	Lca-RTP3-q-2-F	GGCCGTGGCCGGGGGACAGT	61	181
		Lca-RTP3-q-2-R	CGCCATCGTCGTCGTCGTCGTCA		
c12370_g3_i1	heat shock protein 70	Lca-HSP70-q-F	CGTCCCTGATTAAACGGAACACCAC	64	120
		Lca-HSP70-q-R	TGGTCATGGCTCTTTCACCTTCGTA		
c3737_g1_i1	interferon-induced very large gtpase 1-like	Lca-iivlg1-q-F	CAGGGGTGTGTTACTGTCTCCATGT	64	97
		Lca-iivlg1-q-R	TCCAGGTGTGTCTTTGTTTTGGCAG		
c11620_g1_i1	heat shock protein 30-like	Lca-HSP30-q-F	AGCAACACAGAGGAGAACAGACACT	64	146
		Lca-HSP30-q-R	TGCGTACAGGCCAATAGAAGTCCAT		
c7898_g1_i1	viperin	Lca-viperin-q-F	ACAGTTTTGACGAAGCAACCAACCA	64	129
		Lca-viperin-q-R	TGACCGAGTTGATTTTGAAGGCCAC		
c12544_g2_i1	urokinase plasminogen activator surface receptor-like	Lca-upasrl-q-F	GCTCATTTCTGCTCCGATGATTCCC	63	105
		Lca-upasrl-q-R	TCTGTGGGAGGAGGAAAGACAACAG		
c38898_g1_i1	cc chemokine	Lca-ccc-q-F	GCGTGCAGAGATTCTTGCCTTTTCT	64	140
		Lca-ccc-q-R	TATGGACTGCTGCCTGACTGTCAAA		
c144_g1_i1	c-x-c motif chemokine 6-like	Lca-cxcmc-q-F	TACCCGGCAACCATCTTCTGTAACA	64	143
		Lca-cxcmc-q-R	ACTGTGGTGGAGACTTTCTTGCTCT		

Continued Table A3

c65346_g1_i1	interferon-induced protein with tetratricopeptide repeats 5-like	Lca-iipwtr-q-F	ACCTGGCTGTTTACTACCTTGAGCA	64	103
		Lca-iipwtr-q-F	CAGCTGACAGGATTTCTCAGGACCT		
c5512_g2_i1	interferon stimulated protein 15	Lca-isp15-q-F	ATGTCGTAGGTGTTTAACTGCCCCT	64	137
		Lca-isp15-q-R	AACGATTCACAGCCCATCACCTACT		
c5208_g2_i1	interferon alpha-inducible protein 27-like protein 2-like	Lca-iaip27-q-F	TGTGAATGCTCTGTCATGGATTGCC	64	140
		Lca-iaip27-q-R	TCACAGTCATAACACCTCCTGCTCC		
c16863_g1_i1	stimulator of interferon genes	Lca-sig-q-F	TTCCTCTCAACGCCAACATCTCTCA	64	94
		Lca-sig-q-F	GCCCTGTCTATCTCTTTGTTCGGGA		
c12542_g9_i1	interferon regulatory factor 3	Lca-irf3-q-F	CGTGTATCTGACTTCTGACCCCGAA	64	149
		Lca-irf3-q-R	GATACTTGCCGCTGTCAATCTGAGC		
c8372_g1_i1	interferon inducible mx protein	Lca-mx-q-F	CCTGCGCATGTGGAGAAAAAGATT	57	143
		Lca-mx-q-R	GCGATGCCAGGCAAGTCTATGAGT		

Table A4 Primer pairs for cloning and association study in Chapter 4

Primer name	Primer sequence (5' - 3')	Target	Product length (bp)	Tm (°C)
Lca-viperin-Hindiii-F	CCCAAGCTTATGGTCCTCATCTCGTCTCT	*Viperin*	1074	60
Lca-viperin-Hindiii-R	CCCAAGCTTCCAGTCCAGCTTCATGTCA			
Lca-RTP3-2-Hindiii-F	CCCAAGCTTATGGCACACGCAGAGTG	*RTP3*	498	59
Lca-RTP3-2-Hindiii-R	CCCAAGCTTCTCTCTTGTACAGATGCC			
Lca-viperin-SSR1-F	CAGTGTCAGCATTGTCAGCA	(CT)6	244	55
Lca-viperin-SSR1-R	CTGGTTGGTTGCTTCGTCAA			
Lca-RTP3-SSR1-F	GCTCCGGCTGATCAGCAAAATC	(GT)19TT (TG)5	462	55
Lca-RTP3-SSR1-R	TTAAAGCCACAAAATACAAAAACATACATA			
Lca-RTP3-SSR2-F	ACATGCCGAGAGTTGGGTTA	(AAAAG)4	410	55
Lca-RTP3-SSR2-R	GCCGCTACTGTTATTGCTGT			