

## Dataset Paper

# Subtracted Transcriptome Profile of Tiger Shrimp (*Penaeus monodon*) That Survived WSSV Challenge

**Benedict A. Maralit,<sup>1</sup> Minerva Fatimae H. Ventolero,<sup>1</sup> Mary Beth B. Maningas,<sup>2</sup> Edgar C. Amar,<sup>3</sup> and Mudjekeewis D. Santos<sup>1</sup>**

<sup>1</sup> Genetic Fingerprinting Laboratory, National Fisheries Research and Development Institute, Corporate 101 Building, Mother Ignacia Avenue, 1103 Quezon City, Philippines

<sup>2</sup> Department of Biological Science, College of Science/Research Center for the Natural Sciences, Thomas Aquinas Research Center, University of Santo Tomas, España Boulevard, 1008 Manila, Philippines

<sup>3</sup> Aquaculture Department, Southeast Asian Fisheries Development Center (SEAFDEC AQD), Tigbauan Iloilo, Philippines

Correspondence should be addressed to Mudjekeewis D. Santos; [mudjiesantos@yahoo.com](mailto:mudjiesantos@yahoo.com)

Received 15 July 2013; Accepted 5 September 2013; Published 8 May 2014

Academic Editors: A. Hannoufa, Y. Okazaki, and B. Padhuksahasram

Copyright © 2014 Benedict A. Maralit et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There is increased interest in the development of virus-resistant or improved shrimp stock because production is currently hindered by outbreaks and limited understanding of shrimp defense. Recent advancement now allows for high-throughput molecular studies on shrimp immunity. We used next-generation sequencing (NGS) coupled with suppression subtractive hybridization (SSH) to generate a transcriptome database of genes from tiger shrimp that survived White spot syndrome virus (WSSV) challenge. A total of 9,597 unique sequences were uploaded to NCBI Sequence Read Archive with accession number SRR577080. Sixty-five unique sequences, 6% of the total, were homologous to genes of *Penaeus monodon*. Genes that were initially related to bacterial infection and environmental stress such as 14-3-3 gene, heat shock protein 90, and calreticulin were also found including a few full-length gene sequences. Initial analysis of the expression of some genes was done. Hemocyanin, ferritin, and fortilin-binding protein exhibited differential expression between survivor and control tiger shrimps. Furthermore, candidate microsatellite markers for brood stock selection were mined and tested. Four trinucleotide and one dinucleotide microsatellites were successfully amplified. The study highlights the advantage of the NGS platform coupled with SSH in terms of gene discovery and marker generation.

## 1. Introduction

Although a lot of best management practices, detection methods, and disease intervention have been developed, shrimp production is still continuously affected by viral infections. Because of this, greater attention has now been placed on molecular studies for application on genetic breeding programs. As seen in recent years, there is an increase in the number of studies that aim to elucidate the function of several previously unknown genes and their viral response pathways. Expression and characterization studies of one or a group of similar genes are currently gaining pace. Likewise, large databases from which immune-related genes may be determined are now publicly available and are continuously generated.

Some notable genes that have been recently described from Expressed Sequence Tags (EST) databases are hemocyanin, ferritin, and fortilin-binding protein [1–5]. Many subtraction libraries have also been used to associate mined genes from EST to specific roles in immunity. For instance, cathepsins [1, 6] and penaeidins [7–9] were proven to be immune related based on several separate cDNA library studies [10]. Using the same large-scale sequencing approach, genes currently related to bacterial infection and environmental stress, such as heat shock proteins [11], hemocyte homeostasis-associated protein [12], 14-3-3 gene, and calreticulin, are also suggested to play important roles in White spot syndrome virus (WSSV) infection or resistance.

As seen, large-scale transcriptome analysis and EST databases have established many immune related genes, which

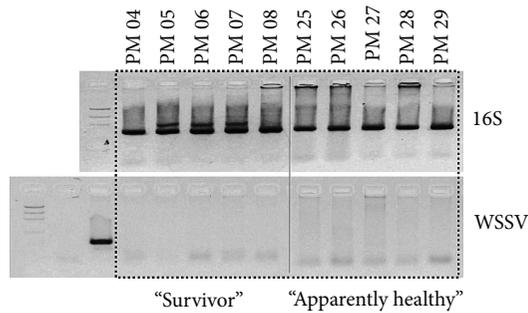


FIGURE 1: PCR-based detection of White spot syndrome virus in all *Penaeus monodon* samples. Top: internal controls; bottom lane 1: ladder; lane 2: negative control; lane 3: positive control; lanes 4–8: survivor samples; lanes 9–13: control shrimp samples.

were further characterized by expression or functional analyses [1, 8, 9, 12, 13]. Indeed, its usefulness has been proven in many aspects of shrimp health. Recent transcriptome approaches such as microarrays and next-generation sequencing (NGS) have also provided efficient applications for rapid determination of quantitative trait loci and genetic markers from immune-related genes [1, 14–16] because these approaches provide longer average sequence reads of 600 base pairs (bp) than traditional ones. The markers generated from these platforms may be used for a lot of genetic studies integral to taxonomy or aquaculture. In fact, the latest breeding technologies in plants make use of next-generation sequencing data for rapid marker development [17]. In shrimp, traditional selective breeding programs have already been useful for producing WSSV resistant *Penaeus vannamei* [1, 18–21]. However, this classical approach entails high-input requirements of capital, time, and research effort. As seen in plants, marker-assisted selection [22], a modified approach in traditional selective breeding, proves successful. It is, thus, the goal of this study to create a database of genes from which candidate markers may be mined and used in marker-assisted selection (MAS) of WSSV-resistant or tolerant *P. monodon*. Suppression subtractive hybridization (SSH) coupled with NGS was used in this study to generate the desired database. We show that this platform can be used to further study genes that may have relevant function in WSSV immunity. Likewise, we identified resistance-related markers such as microsatellites that may be amplified and analyzed for downstream genetic applications such as semiquantitative expression analysis.

## 2. Methodology

Sample collection was as follows. Shrimps for challenge experiments were collected from a farm located in Negros Occidental, Philippines. Shrimp fry used to stock the farm came from a hatchery that practices biosecurity monitoring. Before they were transported to a biosecure facility in SEAFDEC AQD, samples were tested for WSSV and other viruses. Subsequently, they were quarantined in this facility for one month as they were being acclimated to tank conditions.

For WSSV challenge, an *in vivo* titration test by intramuscular injection of the viral inoculum was conducted to determine the dose that resulted in 50% mortality ( $SID_{50}$ ) of test shrimps (155 days old). To do this, gills from moribund shrimp were collected and homogenized in PBS. The homogenate was centrifuged at 5000 rpm. The resulting supernatant was filtered in  $0.45\ \mu\text{m}$  sterile membrane filter, serially diluted ( $10^{-3}$  to  $10^{-9}$ ) in PBS, and  $100\ \mu\text{L}$  was injected into the abdominal muscle region of each shrimp. Mortality was observed for 15 days. The computed  $SID_{50}$  was  $10^{-6.45}\ \text{mL}^{-0.1}$  and the titer of the undiluted supernatant was  $10^{7.45}\ \text{SID}_{50}\ \text{mL}^{-1}$ . The challenge test was conducted by injecting each shrimp with  $1\ \text{SID}_{50}$  of the inoculum following the same protocol used in the *in vivo* titration test. Surviving shrimps from this challenge were kept for genetic analysis. Tissue sampling from surviving shrimps (survivor) was done after 30 days of challenge test termination.

DNA extraction and WSSV detection were as follows. A total of 10 representative *P. monodon* samples were collected. Five shrimps that survived viral challenge experiments (30 days after infection) were labeled as “survivor” while five apparently healthy shrimps (no gross signs of disease) were labeled as “control.” Extraction of DNA from gills was performed using CTAB extraction buffer as described by Santos et al. [23]. Spectrophotometry of DNA extracts was performed before proceeding to amplification (PCR) experiments. A total of  $10\ \mu\text{L}$  of reaction mixture was used containing  $2\ \mu\text{L}$  of DNA template,  $0.2\ \text{mM}$  dNTP,  $2\ \mu\text{M}$  of each primer [24, 25], and  $1\times$  buffer already containing  $1.5\ \text{mM}$   $\text{MgCl}_2$  and amplified under the following conditions: initial denaturation at  $95^\circ\text{C}$  for 5 minutes, followed by 30 cycles of denaturation at  $95^\circ\text{C}$  for 30 seconds, annealing at  $52^\circ\text{C}$  for 30 seconds, and extension at  $72^\circ\text{C}$  for another 30 seconds and a final extension for 5 minutes and storage at  $4^\circ\text{C}$ . All DNA extracts were confirmed to contain DNA by PCR amplification using universal 16Sar primers [26]. Based on the PCR detection method described, none of the samples was positive for WSSV (Figure 1).

Library preparation was as follows. Suppression subtractive hybridization (SSH) is one of the most powerful and popular methods for generating subtracted cDNA libraries because the probability of obtaining low-abundance differentially expressed cDNA fragments is very high. It is a suppression PCR technique that uses normalization and subtraction in a single procedure. It was used in this study because downstream analyses, such as filtering after next-generation sequencing of the subtracted library, are simplified in this platform.

RNA extracted from gills using Nucleospin RNA II kit (Machery-Nagel) was subjected to spectrophotometry for the determination of purity and concentration. cDNA was synthesized from the RNA extracts using Clontech SMARTer cDNA synthesis kit. Immediately after, cDNA was amplified using Clontech Advantage II PCR kit through long distance-PCR (ld-PCR) following the recommended protocol. Incubation time was increased from 3 minutes to 6 minutes to ensure synthesis and amplification of longer cDNA transcripts. The cDNA templates with highest quality were used in the subtraction experiment.

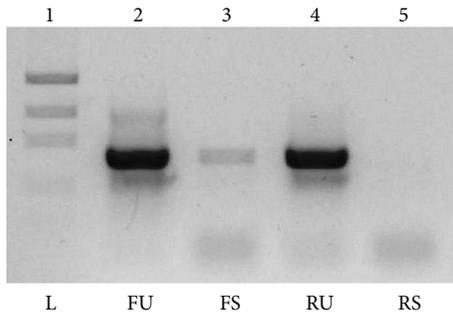


FIGURE 2: Evaluation of the subtraction efficiency of the subtracted cDNAs. PCRs with  $\beta$ -actin-specific primers were performed using forward (lanes 2 and 3) and reverse (lanes 4 and 5) SSH products. Lane 1: ladder; lanes 2 and 4: un-subtracted; lanes 3 and 5: subtracted.

cDNA subtraction was carried out using Clonetechn PCR-Select cDNA Subtraction Kit in accordance with the manufacturer's protocol. For forward subtraction, cDNA synthesized from shrimp sample that survived viral challenge (survivor) was used as driver while cDNA from control shrimp was used as tester. Forward and reverse SSH reactions were performed simultaneously. Tester and driver cDNA populations were briefly digested with *RsaI* and ligated each with one of the two adaptors (adaptor 1 and adaptor 2R) included in the kit. After denaturation at 98°C for 90 s, tester cDNAs were ligated with the adaptor 1 and adaptor 2R separately and then hybridized with an excess of driver cDNAs at 68°C for 8 h. After the first hybridization, the two products (adaptor 1 and adaptor 2R) were mixed together and immediately hybridized again with an excess of freshly denatured driver cDNA for 20 h at 68°C. Differentially expressed products were then selectively amplified by two rounds of PCR using Clonetechn Advantage II PCR kit with primers specific to the two adaptors, resulting in suppression of genes common to the two cDNA populations and enrichment of the differentially expressed genes. The efficiency of the subtraction was finally evaluated by comparing the abundance of the constitutively expressed-actin in the subtracted and un-subtracted products by PCR as described by Zhao et al. [1]. Figure 2 shows that, after subtraction, the constitutive abundance of  $\beta$ -actin was efficiently reduced after SSH.

Next-generation sequencing and bioinformatics were as follows. The forward subtracted product from SSH was sent to Macrogen Inc., Korea (<http://www.macrogen.com/eng/>), for 454 sequencing of 1/4 plate. De novo assembly of the acquired reads was conducted using GS De Novo Assembler v 2.6 (Newbler) with cDNA/transcriptome option. Newbler parameters such as base calling, quality calling, and assembly quality filters were set to default. Singletons were cleaned by SeqClean (<http://sourceforge.net/projects/seq-clean/>) and further assessed and controlled by Lucy (<http://lucy.sourceforge.net/>) with minimum length filter of 100 bp. A complete workflow of the study is exhibited in Figure 3.

As seen in Table 2, a total of 76,889,928 bases, which comprised 240,616 reads, were generated from NGS. Approximately 52% of the reads were assembled to form 125,598 sequences with an average read length of 319 ranging from

100 to 600 base pairs. De novo assembly generated 4,090 contigs with N50 of 504. Likewise, 9,526 singletons were produced leaving out 10,108 outliers and 2,650 too short sequences. Two phases of sequence cleaning using SeqClean and Lucy generated 5,507 singletons, completing a total of 9,597. This number of unique sequences is comparable to those stored in *Penaeus monodon* Genome Database (<http://sysbio.iis.sinica.edu.tw/page/index.php>) and the *Penaeus monodon* gene discovery project [27]. The genome size of tiger shrimp is not yet fully determined but the 1C nuclear DNA content of *P. monodon* was estimated to be ~72.2% of the human genome, that is, ~2.53 pg DNA per nucleus or  $2.17 \times 10^9$  bp per haploid genome [28]. In Table 2, the first part presents Reads Status where the number of reads counts the reads used in the assembly computation; number of bases, the read's bases used in the assembly computation; assembled, the number of reads that are fully incorporated into the assembly; partial, the number of reads with only a part being included in the assembly; singleton, the number of reads that did not overlap with any other reads in the input; repeat, the number of reads deemed to be from repeat regions; outlier, the number of reads that were identified by the GS De Novo Assembler as problematic; and too short, the number of reads that were too short to be used in the computation. The second part of the table presents the frequency information on isogroups or collection of contigs containing reads that imply connections between them. The number of isogroups corresponds to the number of isogroups identified; average contig count, the average count of contigs in the isogroups; largest contig count, the largest count of contigs in the isogroups; number with one contig, the number of isogroups assembled by one contig; average isotig count, the average count of isotigs in the isogroups; largest isotig count, the largest count of isotigs in the isogroups; number with one isotig, the number of isogroups assembled by one isotig. The third part presents the frequency information of isotigs (analogous to an individual transcript). The number of isotigs correspond to the number of contigs identified; average contig count, the average count of contigs in the isotigs; largest contig count, the largest count of contigs in the isotigs; number with one contig, the number of isotigs assembled by one contig; number of bases, the total number of bases in the isotig; average isotig size, the average isotig size; N50 isotig size, the N50 isotig size (an N50 means that half of all bases reside in isotigs of this size or longer); largest isotig size, the size of the largest isotig. The last part in the table presents the frequency information of singletons. The number of singletons corresponds to the number of singletons identified; valid, the number of valid singletons after SeqClean; short, the number of sequences less than 100 bp; LowQual, undetermined base >3% in clear range; dust, standalone by low complexity (dust) filter <40 nt; ShortQ, trimmed sequence less than 100 bp; valid (after second cleaning), the number of valid singletons after Lucy; trashed, the number of low-quality sequences.

Processed reads and generated contigs were aligned to the sequences publicly available in the NCBI GenBank database. Separate BLAST analyses were done for isotigs (Dataset Item 1 (Table)) and singletons (Dataset Item 2 (Table)). Of the sequences generated from NGS, 598 (14.62%) had significant

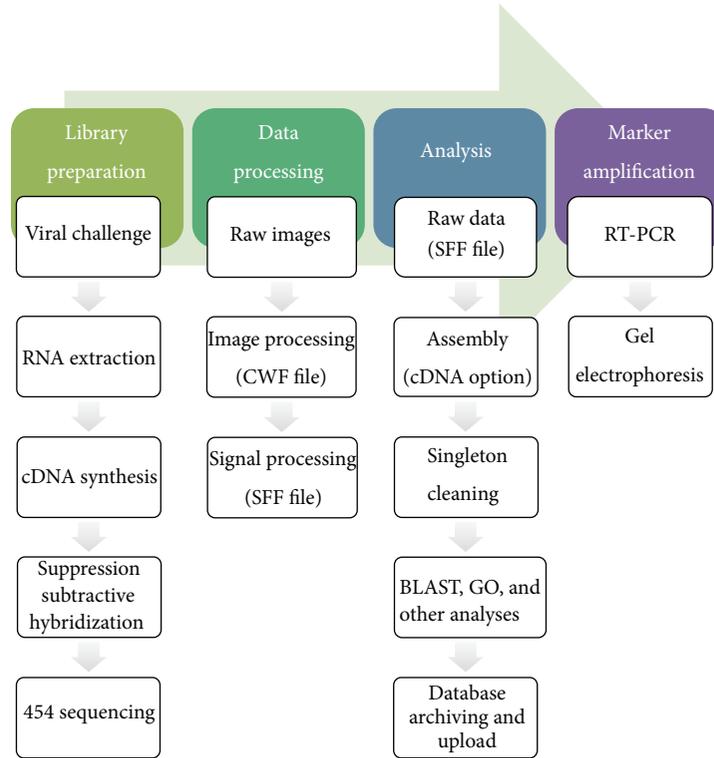


FIGURE 3: Experiment workflow.

matches ( $\leq 1.0E - 3$ ) to the nonredundant (nr) nucleotide database of GenBank after BLAST analysis with default options. Twenty-six (26) sequences (4.25%) of those with BLASTn hits had  $E$ -values of  $\leq 1.0E - 100$  and, thus, were considered to have highly significant homology [29]. About 296 sequences (49.50%) had homology values between  $1.0E - 99$  and  $1.0E - 20$ , which were considered as moderately similar to the sequences in GenBank. The remaining 276 isotigs (46.15%) had weak homologies or similarities ( $E$ -values between  $1.0E - 19$  and  $1.0E - 3$ ). Meanwhile, out of the 5,507 cleaned singletons, 414 had significant hits to GenBank sequences. One hundred sixty-one (161) (38.89%) of those annotated had moderate similarity to the gene sequences in GenBank. Out of the combined 1,012 BLASTn homologous sequences, 9% were similar to *Daphnia pulex* nucleotide sequences (Figure 4).

The remaining majority had significant homologies to several shrimp species such as *Penaeus monodon* (6%), *Litopenaeus vannamei* (3%), *Marsupenaeus japonicus* (3%), and *Fenneropenaeus chinensis* (2%). Dataset Item 3 (Table) lists the database sequences with homology to *P. monodon* genes in the GenBank nucleotide database. Resulting gene ontology (GO) assignment showed that 88% of the total isotigs (Dataset Item 4 (Table)) and 94% of the singletons (Dataset Item 5 (Table)) had no significant hits to the Genbank GO database. Tables 3 and 4 show the isotigs and singletons GO analysis results, respectively. 488 isotigs and 336 singletons were assigned and distributed to different categories based on gene ontology. Complete lists of isotig and singleton BLASTx or GO assignments are exhibited in

Dataset Item 6 (Table) and Dataset Item 7 (Table), respectively.

In terms of similarities to model organisms, the majority of the 824 GO assigned sequences (40%) had significant hits to genes of *Drosophila melanogaster*. Genes from *Danio rerio* also had homologies to 23% of the sequences. The remaining sequences were significantly homologous to the genes of *Rattus norvegicus* (14%), *Mus musculus* (10%), and other species (Figure 5).

Microsatellite mining was as follows. Using msatcommander with search options for trinucleotide repeats or higher, 25 microsatellite markers were identified in Dataset Item 8 (Table). Five (20%) of these have BLASTn matches. The homologous genes, however, were putative or predicted. Some of these were amplified using RT-PCR with the following mixture:  $\mu\text{L}$  1 cDNA template in 1x PCR buffer with 2 mM  $\text{MgCl}_2$ , 0.2 mM dNTPs, 0.5  $\mu\text{M}$  forward and reverse primers, 1 unit Taq, and distilled water. The PCR mix was run in a thermal cycler with the following conditions: 95°C for 5 minutes, 30 cycles of 95°C for 30 seconds, 52°C for 30 seconds, 72°C for 30 seconds, and a final extension of 72°C for 5 minutes. Successfully amplified microsatellite markers are shown in Figure 6.

The successful amplification of these microsatellites indicates that repetitive segments may also be found in the transcriptome and they may be used as genetic markers relating to desired phenotype. Once confirmed to be polymorphic and after further field testing using genomic DNA, these microsatellite markers may eventually be established as markers for WSSV resistance in shrimp.

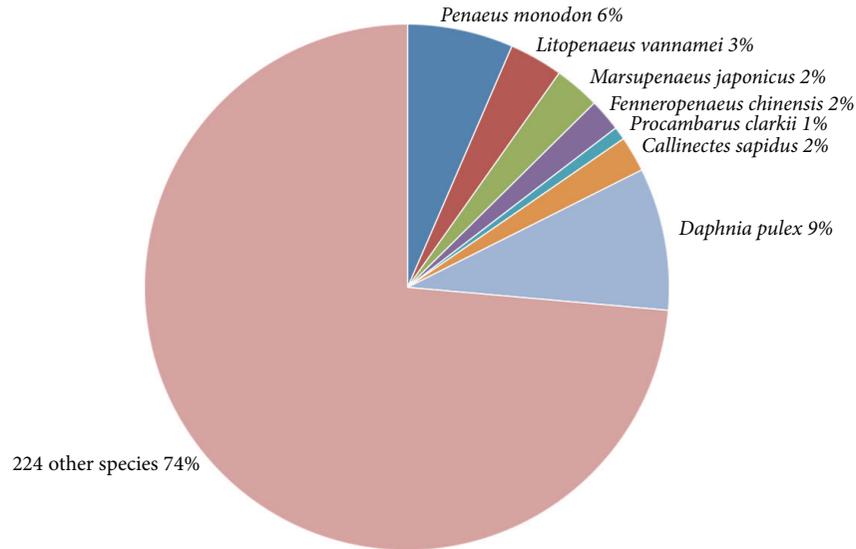


FIGURE 4: Organisms with most nucleotide hits: the proportion of unique sequences with BLASTn matches ( $E\text{-value} \leq 10^{-3}$ ) in the nonredundant nucleotide database classified according to the organism with the “top hit” nucleotide sequence.

Semiquantitative expression analysis was as follows. The concentration and purity of the RNA extracts were determined by spectrophotometry prior to cDNA synthesis and RT-PCR. All extracts were found to contain quality RNA (OD 260/280 of 1.8 to 2.0). Gene expression was analyzed using semiquantitative PCR. PCR mix was prepared using the same concentrations as described in microsatellite amplification. Running conditions, however, were modified into 25 cycles with a 50°C cycle annealing temperature. After the run, 2  $\mu\text{L}$  of each amplicon was mixed with 2  $\mu\text{L}$  of loading dye and then electrophoresed in 1% agarose gel. Relative expression was determined based on the fluorescence of 2  $\mu\text{L}$  DNA ladder loaded on the same gel and computed using BioDocAnalyze Software bundled with the Biometra Gel Documentation system. Ferritin and fortilin-binding protein, whose full-length coding regions were found in the library, were successfully amplified. Differential expression of these genes was clearly observed in the semiquantitative PCR profiles (Figure 7).

$\beta$ -Actin was used as negative control [1] while hemocyanin, a known differentially expressed gene in virus resistant shrimps [13], was used as a positive control. It can be seen that fortilin-binding protein is significantly upregulated in “survivor” shrimps. Primers and annealing temperatures used were listed in Table 1.

### 3. Dataset Description

The dataset associated with this Dataset Paper consists of 8 items which are described as follows.

*Dataset Item 1 (Table).* A list of isotigs and BLAST results. The column Query Name presents the isotig number designated by the assembly program; Query Length, the length of isotig sequence in base pairs; Query Start, the start of alignment in query; Query End, the end of alignment in query; Query Coverage, the percentage of query covered

by alignment to the database sequence; Subject Description, the description or title of the matched database sequence; Subject Accession, the accession of the matched database sequence; Subject Length, the length of the matched database sequence in base pairs; Subject Start, the start of alignment in subject; Subject End, the end of alignment in subject; Subject Coverage, the percentage of database sequence covered by alignment to the query sequence; Bit Score, the alignment bit score;  $E$ -Value Score, the alignment expect value from the database sequence; Match/Total Identities, the number of identical matches; Identities Percentage, the percentage of identical matches; Match/Total Positives, the number of positive scoring matches; Positives Percentage, the percentage of positive scoring matches; Match/Total Gaps, the number of gaps in the alignment; Gaps Percentage, the percentage of gaps in the alignment; Query Frame, the reading frame of the query sequence.

*Column 1:* Query Name

*Column 2:* Query Length

*Column 3:* Query Start

*Column 4:* Query End

*Column 5:* Query Coverage

*Column 6:* Subject Description

*Column 7:* Subject Accession

*Column 8:* Subject Length

*Column 9:* Subject Start

*Column 10:* Subject End

*Column 11:* Subject Coverage

*Column 12:* Score Bit

*Column 13:*  $E$ -Value Score

*Column 14:* Match/Total Identities

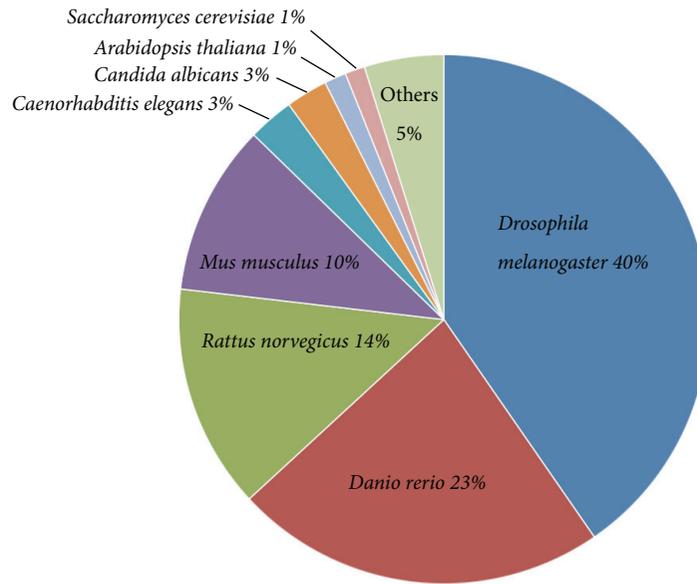


FIGURE 5: Organisms with most protein hits: the proportion of putative transcripts with BLASTx matches ( $E\text{-value} \leq 10^{-3}$ ) in the nonredundant protein database classified according to the organism with the “top hit” protein sequence.

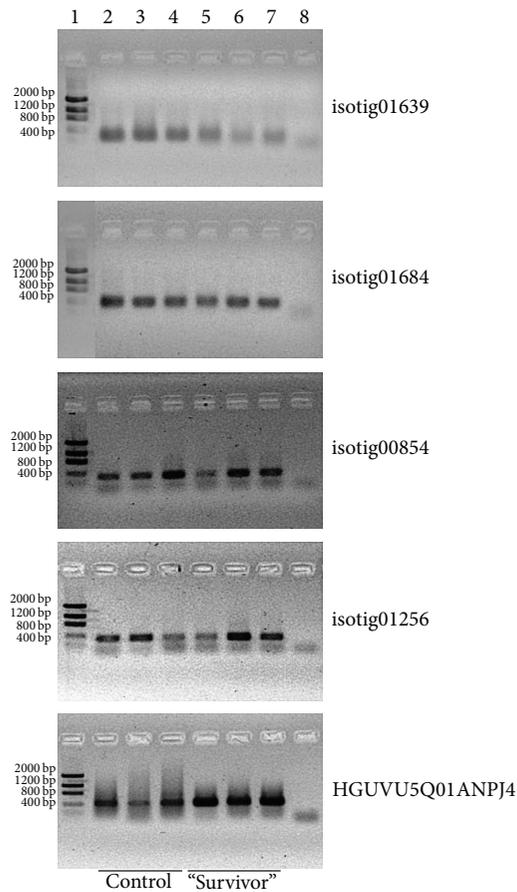


FIGURE 6: Successful amplification of microsatellites in *P. monodon* samples. Leftmost: marker lengths; lane 1: ladder; lanes 2 to 4: control shrimps; lanes 5 to 7: “survivor” shrimps; lanes 8: negative control; rightmost: microsatellite identifier.

TABLE 1: Primers used in RT-PCR.

Product	Identifier	5' to 3' sequence	Melting temp.
170	isotig001639	Forward: TCAAAGTTAGCTCCTAGACCA	52
		Reverse: AGCTCCTACTCCTGTTTCTGT	55
146	isotig001684	Forward: GCCTCCTATAAACGCTAAAAT	50
		Reverse: ATTACCCAGAATTCACCTGGTT	51
325	isotig00854	Forward: CTGCAGGTCATGCTCAACG	60
		Reverse: AGCTCCCACACATCAAGGG	60
314	isotig01256	Forward: ACGACTCGAAGGAAGGCTC	60
		Reverse: TCCCGAATGTCCGTTCTCC	60
154	HGUVU5Q01ANPJ4	Forward: TCGCGACGAGTACATGGG	60
		Reverse: GTCTTCACTGTATTCAATTGGCTG	60
268	Hemocyanin	Forward: ACAGTCTCCCTCCATACACCGTGG	63
		Reverse: AGGCCAGGCGAAAACGCGAA	64
279	Ferritin	Forward: GCCCGGGCAGGTACTCTCCA	66
		Reverse: TGGCCGCATCGTTCTCCAGC	64
223	Fortilin-binding protein	Forward: GCTCGCAGTTAACCTCCTCTGCC	63
		Reverse: GGCGTCCACCGTGCAAGAAGT	64

TABLE 2: A list of database reads, sequences, and other components with their corresponding frequencies.

Number of reads	Number of bases	Assembled	Partial	Singleton	Repeat	Outlier	Too short
240,616	76,099,104	125,598	91,137	9,526	1,597	10,108	2,650
Number of isogroups	Average contig count	Largest contig count	Number with one contig	Average isotig count	Largest isotig count	Number with one isotig	
3,724	2	2,813	3,573	1	13	3,577	
Number of isotigs	Average contig count	Largest contig count	Number with one contig	Number of bases	Average isotig size	N50 isotig size	Largest isotig size
4,090	1	6	3,732	1,635,969	400	504	1,921
Number of singletons	Valid	Short	LowQual	Dust	ShortQ	Valid (after 2nd cleaning)	Trashed
9,526	5,771	3,489	1	73	192	5,507	264

TABLE 3: Isotigs GO analysis result.

GO analysis result	
Biological process	168.6118225
Cellular component	127.6988942
Molecular function	191.6892833
No hits	3602

TABLE 4: Singletons GO analysis result.

GO analysis result	
Biological process	101.4409221
Cellular component	93.16145113
Molecular function	141.3976268
No hits	5171

Column 15: Identities Percentage (%)

Column 16: Match/Total Positives

Column 17: Positives Percentage (%)

Column 18: Match/Total Gaps

Column 19: Gaps Percentage (%)

Column 20: Query Frame

*Dataset Item 2 (Table).* A list of singletons and BLAST results. The column Query Name presents the singleton code designated by the assembly program; Query Length, the

length of isotig sequence in base pairs; Query Start, the start of alignment in query; Query End, the end of alignment in query; Query Coverage, the percentage of query covered by alignment to the database sequence; Subject Description, the description or title of the matched database sequence; Subject Accession, the accession of the matched database sequence; Subject Length, the length of the matched database sequence in base pairs; Subject Start, the start of alignment in subject; Subject End, the end of alignment in subject; Subject Coverage, the percentage of database sequence covered by alignment to the query sequence; Bit Score, the alignment

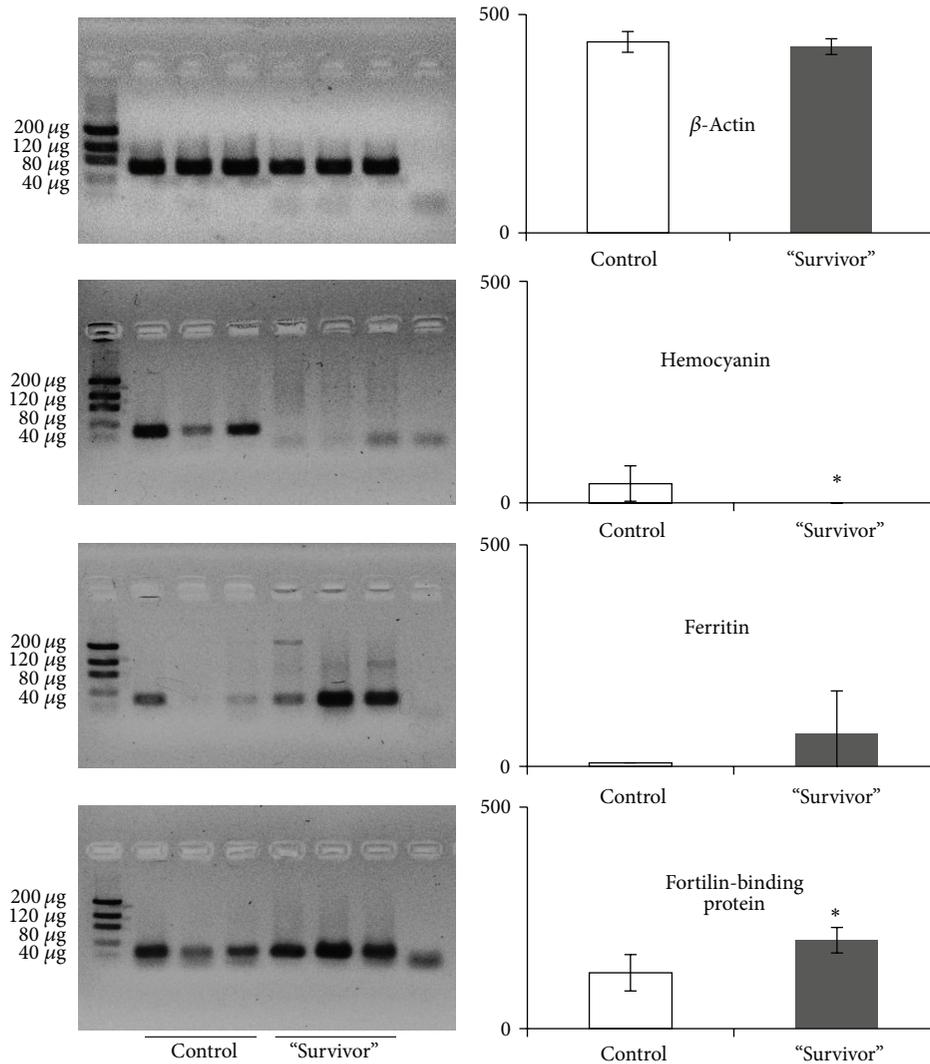


FIGURE 7: Successful amplification of immune-related genes in *P. monodon* samples. Lane 1: ladder; lanes 2 to 4: control shrimps; lanes 5 to 7: "survivor" shrimps; lane 8: negative control. Asterisk indicates significant difference between  $\beta$ -actin normalized fluorescence scores of control and "survivor" shrimps based on *t*-test at 95% confidence interval.

bit score; *E*-Value Score, the alignment expect value from the database sequence; Match/Total Identities, the number of identical matches; Identities Percentage, the percentage of identical matches; Match/Total Positives, the number of positive scoring matches; Positives Percentage, the percentage of positive scoring matches; Match/Total Gaps, the number of gaps in the alignment; Gaps Percentage, the percentage of gaps in the alignment; Query Frame, the reading frame of the query sequence.

Column 1: Query Name  
 Column 2: Query Length  
 Column 3: Query Start  
 Column 4: Query End  
 Column 5: Query Coverage  
 Column 6: Subject Description

Column 7: Subject Accession  
 Column 8: Subject Length  
 Column 9: Subject Start  
 Column 10: Subject End  
 Column 11: Subject Coverage  
 Column 12: Score Bit  
 Column 13: *E*-Value Score  
 Column 14: Match/Total Identities  
 Column 15: Identities Percentage (%)  
 Column 16: Match/Total Positives  
 Column 17: Positives Percentage (%)  
 Column 18: Match/Total Gaps  
 Column 19: Gaps Percentage (%)  
 Column 20: Query Frame

*Dataset Item 3 (Table).* A list of dataset sequence names (Name), their corresponding annotation according to the Genbank nonredundant nucleotide database (NR Annotation), and related organisms (Related Organism).

*Column 1:* Name

*Column 2:* NR Annotation

*Column 3:* Related Organism

*Dataset Item 4 (Table).* A list of isotig frequencies of major gene ontology (GO) categories and subcategories. Since a gene product could be assigned to more than one GO term, the percentages in each main category do not add up to exactly 100 percent. Supplementary information is included where the isotig number, isogroup number, length, and number of contigs comprising the contig are listed per GO term.

*Column 1:* GO Category

*Column 2:* GO Subcategory

*Column 3:* Frequency Percentage (%)

*Column 4:* Supplementary Information

*Dataset Item 5 (Table).* A list of singleton frequencies of major gene ontology (GO) categories and subcategories. Since a gene product could be assigned to more than one GO term, the percentages in each main category do not add up to exactly 100 percent. Supplementary information is also included where the singleton code is listed per GO term.

*Column 1:* GO Category

*Column 2:* GO Subcategory

*Column 3:* Frequency Percentage (%)

*Column 4:* Supplementary Information

*Dataset Item 6 (Table).* A list of isotigs and BLAST2GO results. The column Query Name presents the isotig number designated by the assembly program; Query Length, the length of isotig sequence in base pairs; Query Start, the start of alignment in query; Query End, the end of alignment in query; Query Coverage, the percentage of query covered by alignment to the database sequence; Subject Description, the description or title of the matched database sequence with specific details that contains hit ID, sequence symbol, species number and name, GO number, specific GO term, and evidence source; Subject Length, the length of the matched database sequence in base pairs; Subject Start, the start of alignment in subject; Subject End, the end of alignment in subject; Subject Coverage, the percentage of database sequence covered by alignment to the query sequence; Bit Score, the alignment bit score; *E*-Value Score, the alignment expect value from the database sequence; Match/Total Identities, the number of identical matches; Identities Percentage, the percentage of identical matches; Match/Total Positives, the number of positive scoring matches; Positives Percentage, the percentage of positive scoring matches; Match/Total

Gaps, the number of gaps in the alignment; Gaps Percentage, the percentage of gaps in the alignment; Query Frame, the reading frame of the query sequence.

*Column 1:* Query Name

*Column 2:* Query Length

*Column 3:* Query Start

*Column 4:* Query End

*Column 5:* Query Coverage

*Column 6:* Subject Description

*Column 7:* Subject Length

*Column 8:* Subject Start

*Column 9:* Subject End

*Column 10:* Subject Coverage

*Column 11:* Score Bit

*Column 12:* *E*-Value Score

*Column 13:* Match/Total Identities

*Column 14:* Identities Percentage (%)

*Column 15:* Match/Total Positives

*Column 16:* Positives Percentage (%)

*Column 17:* Match/Total Gaps

*Column 18:* Gaps Percentage (%)

*Column 19:* Query Frame

*Dataset Item 7 (Table).* A list of singletons and BLAST2GO results. The column Query Name presents the singleton code designated by the assembly program; Query Length, the length of isotig sequence in base pairs; Query Start, the start of alignment in query; Query End, the end of alignment in query; Query Coverage, the percentage of query covered by alignment to the database sequence; Subject Description, the description or title of the matched database sequence with specific details that contains hit ID, sequence symbol, species number and name, GO number, specific GO term, and evidence source; Subject Length, the length of the matched database sequence in base pairs; Subject Start, the start of alignment in subject; Subject End, the end of alignment in subject; Subject Coverage, the percentage of database sequence covered by alignment to the query sequence; Bit Score, the alignment bit score; *E*-Value Score, the alignment expect value from the database sequence; Match/Total Identities, the number of identical matches; Identities Percentage, the percentage of identical matches; Match/Total Positives, the number of positive scoring matches; Positives Percentage, the percentage of positive scoring matches; Match/Total Gaps, the number of gaps in the alignment; Gaps Percentage, the percentage of gaps in the alignment; Query Frame, the reading frame of the query sequence.

*Column 1:* Query Name

*Column 2:* Query Length

*Column 3:* Query Start

*Column 4:* Query End

Column 5: Query Coverage  
 Column 6: Subject Description  
 Column 7: Subject Length  
 Column 8: Subject Start  
 Column 9: Subject End  
 Column 10: Subject Coverage  
 Column 11: Score Bit  
 Column 12: E-Value Score  
 Column 13: Match/Total Identities  
 Column 14: Identities Percentage (%)  
 Column 15: Match/Total Positives  
 Column 16: Positives Percentage (%)  
 Column 17: Match/Total Gaps  
 Column 18: Gaps Percentage (%)  
 Column 19: Query Frame

*Dataset Item 8 (Table)*. A list dataset sequences with microsatellites. The column Sequence identifier presents the isotig number or singleton code; Length, the sequence length in base pairs; Repeats, the number and motif of microsatellite repeats; Left Start Primer, the position of forward primer along the sequence; Right Start Primer, the position of reverse primer along the sequence; Blastn Match, the annotation based on Blastn.

Column 1: Sequence Identifier  
 Column 2: Length  
 Column 3: Repeats  
 Column 4: Left Start Primer  
 Column 5: Right Start Primer  
 Column 6: Blastn Match

#### 4. Concluding Remarks

The transcriptome assembly described by this paper is to be used in conjunction with Dataset Items 1–8 (Tables) to facilitate easy and rapid identification of any gene of interest from the “survivor” tiger shrimp transcriptome. This dataset represents a large pool of sequences that needs to be further characterized for additional understanding of WSSV *P. monodon* interaction. Once validated, candidate markers mined from the library including the generated microsatellites may be used for genetic linkage mapping.

#### Dataset Availability

The dataset associated with this Dataset Paper is dedicated to the public domain using the CC0 waiver and is available at <http://dx.doi.org/10.1155/2014/807806/dataset>. In addition, the sequence database has been uploaded to NCBI Sequence Read Archive with accession number SRR577080 (<http://www.ncbi.nlm.nih.gov/sra>).

#### Conflict of Interests

The authors declare that there is no conflict of interests.

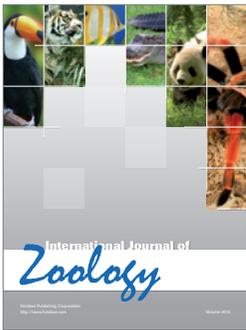
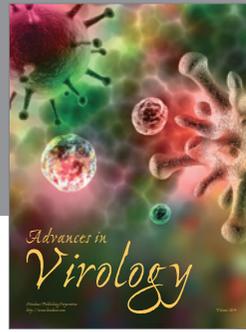
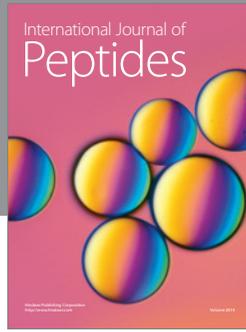
#### Acknowledgments

The authors would like to thank Dr. Adelaida T. Calpe of the Philippine Council for Agriculture, Aquatic and Natural Resources Research and Development (PCAARRD) for guiding the project’s implementation; Mr. Kenneth Kim; the Biochemistry Laboratory of the Thomas Aquinas Research Complex; Roselyn Aguila, Sweedy Kay Perez, and the rest of the staff of Genetic Fingerprinting Laboratory; and the NFRDI administration.

#### References

- [1] Z.-Y. Zhao, Z.-X. Yin, S.-P. Weng et al., “Profiling of differentially expressed genes in hepatopancreas of white spot syndrome virus-resistant shrimp (*Litopenaeus vannamei*) by suppression subtractive hybridisation,” *Fish and Shellfish Immunology*, vol. 22, no. 5, pp. 520–534, 2007.
- [2] D. Pan, N. He, Z. Yang, H. Liu, and X. Xu, “Differential gene expression profile in hepatopancreas of WSSV-resistant shrimp (*Penaeus japonicus*) by suppression subtractive hybridization,” *Developmental and Comparative Immunology*, vol. 29, no. 2, pp. 103–112, 2005.
- [3] M. Tonganunt, B. Nupan, M. Saengsakda et al., “The role of Pmfortilin in protecting shrimp from white spot syndrome virus (WSSV) infection,” *Fish and Shellfish Immunology*, vol. 25, no. 5, pp. 633–637, 2008.
- [4] T. Panrat, P. Sinthujaroen, B. Nupan, W. Wanna, M. T. Tammi, and A. Phongdara, “Characterization of a novel binding protein for Fortilin/TCTP—component of a defense mechanism against viral infection in *Penaeus monodon*,” *PLoS ONE*, vol. 7, no. 3, Article ID e33291, 2012.
- [5] A. Clavero-Salas, R. R. Sotelo-Mundo, T. Gollas-Galván et al., “Transcriptome analysis of gills from the white shrimp *Litopenaeus vannamei* infected with White Spot Syndrome Virus,” *Fish and Shellfish Immunology*, vol. 23, no. 2, pp. 459–472, 2007.
- [6] J. Robalino, J. S. Almeida, D. McKillen et al., “Insights into the immune transcriptome of the shrimp *Litopenaeus vannamei*: tissue-specific expression profiles and transcriptomic responses to immune challenge,” *Physiological Genomics*, vol. 29, no. 1, pp. 44–56, 2007.
- [7] B. J. Cuthbertson, L. J. Deterding, J. G. Williams et al., “Diversity in penaeidin antimicrobial peptide form and function,” *Developmental and Comparative Immunology*, vol. 32, no. 3, pp. 167–181, 2008.
- [8] J. C. García, A. Reyes, M. Salazar, and C. B. Granja, “Differential gene expression in White Spot Syndrome Virus (WSSV)-infected naïve and previously challenged Pacific white shrimp *Penaeus (Litopenaeus) vannamei*,” *Aquaculture*, vol. 289, no. 3–4, pp. 253–258, 2009.
- [9] E. de la Vega, B. M. Degnan, M. R. Hall, and K. J. Wilson, “Differential expression of immune-related genes and transposable elements in black tiger shrimp (*Penaeus monodon*) exposed to a range of environmental stressors,” *Fish and Shellfish Immunology*, vol. 23, no. 5, pp. 1072–1088, 2007.

- [10] S. Pongsomboon, S. Tang, S. Boonda, T. Aoki, I. Hirono, and A. Tassanakajon, "A cDNA microarray approach for analyzing transcriptional changes in *Penaeus monodon* after infection by pathogens," *Fish and Shellfish Immunology*, vol. 30, no. 1, pp. 439–446, 2011.
- [11] T. Aoki, H.-C. Wang, S. Unajak, M. D. Santos, H. Kondo, and I. Hirono, "Microarray analyses of shrimp immune responses," *Marine Biotechnology*, vol. 13, no. 4, pp. 629–638, 2011.
- [12] A. Prapavorarat, S. Pongsomboon, and A. Tassanakajon, "Identification of genes expressed in response to yellow head virus infection in the black tiger shrimp, *Penaeus monodon*, by suppression subtractive hybridization," *Developmental and Comparative Immunology*, vol. 34, no. 6, pp. 611–617, 2010.
- [13] S. Nayak, S. K. Singh, N. Ramaiah, and R. A. Sreepada, "Identification of upregulated immune-related genes in *Vibrio harveyi* challenged *Penaeus monodon* postlarvae," *Fish and Shellfish Immunology*, vol. 29, no. 3, pp. 544–549, 2010.
- [14] J. Zhang, R. Chiodini, A. Badr, and G. Zhang, "The impact of next-generation sequencing on genomics," *Journal of Genetics and Genomics*, vol. 38, no. 3, pp. 95–109, 2011.
- [15] F. Bontems, L. Baerlocher, S. Mehenni, I. Bahechar, L. Farinelli, and R. Dosch, "Efficient mutation identification in zebrafish by microarray capturing and next generation sequencing," *Biochemical and Biophysical Research Communications*, vol. 405, no. 3, pp. 373–376, 2011.
- [16] S. G. Bavykin, V. M. Mikhailovich, V. M. Zakharyev et al., "Discrimination of *Bacillus anthracis* and closely related microorganisms by analysis of 16S and 23S rRNA with oligonucleotide microarray," *Chemico-Biological Interactions*, vol. 171, no. 2, pp. 212–235, 2008.
- [17] H. Yang, Y. Tao, Z. Zheng, C. Li, M. W. Sweetingham, and J. G. Howieson, "Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L.," *BMC Genomics*, vol. 13, p. 318, 2012.
- [18] J. Cuéllar-Anjel, R. Chamoro, B. White-Noble, P. Schofield, and D. Lightner, "Testing finds resistance to WSSV in shrimp from Panamanian breeding program," *Global Aquaculture Advocate*, pp. 65–66, 2011.
- [19] Z. Huang, Z. Yin, S. Weng, J. He, and S. Li, "Selective breeding and preliminary commercial performance of *Penaeus vannamei* for resistance to white spot syndrome virus (WSSV)," *Aquaculture*, vol. 364–365, pp. 111–117, 2012.
- [20] Y. Huang, A. Hua-shui, Z. Yin et al., "Studies on WSSV resistant and immune characteristics of the 4th generation selective breeding families for resistance to the white syndrome virus (WSSV) of *Litopenaeus vannamei*," *Journal of Fisheries of China*, vol. 34, no. 10, pp. 1550–1558, 2010.
- [21] T. Gitterle, R. Salte, B. Gjerde et al., "Genetic (co)variation in resistance to White Spot Syndrome Virus (WSSV) and harvest weight in *Penaeus* (*Litopenaeus vannamei*)," *Aquaculture*, vol. 246, no. 1–4, pp. 139–149, 2005.
- [22] X.-S. Hu, "A general framework for marker-assisted selection," *Theoretical Population Biology*, vol. 71, no. 4, pp. 524–542, 2007.
- [23] M. D. Santos, G. V. Lopez, and N. C. Barut, "A pilot study on the genetic variation of eastern little tuna (*Euthynnus affinis*) in Southeast Asia," *Philippine Journal of Science*, vol. 139, no. 1, pp. 43–50, 2010.
- [24] B. A. Maralit, C. M. A. Caipang, M. D. Santos, and M. B. B. Maningas, "PCR detection of white spot syndrome virus (WSSV) from farmed pacific white shrimp (*Litopenaeus vannamei*) in selected sites of the Philippines," *AACL Bioflux*, vol. 4, no. 4, pp. 474–480, 2011.
- [25] L. M. Tapay, E. C. B. Nadala Jr., and P. C. Loh, "A polymerase chain reaction protocol for the detection of various geographical isolates of white spot virus," *Journal of Virological Methods*, vol. 82, no. 1, pp. 39–43, 1999.
- [26] A. Meyer, J. M. Morrissey, and M. Schartl, "Recurrent origin of a sexually selected trait in Xiphophorus fishes inferred from a molecular phylogeny," *Nature*, vol. 368, no. 6471, pp. 539–542, 1994.
- [27] A. Tassanakajon, S. Klinbunga, N. Paunglarp et al., "*Penaeus monodon* gene discovery project: the generation of an EST collection and establishment of a database," *Gene*, vol. 384, no. 1–2, pp. 104–112, 2006.
- [28] S.-W. Huang, Y.-Y. Lin, E.-M. You et al., "Fosmid library end sequencing reveals a rarely known genome structure of marine shrimp *Penaeus monodon*," *BMC Genomics*, vol. 12, article 242, 2011.
- [29] J. Quilang, S. Wang, P. Li et al., "Generation and analysis of ESTs from the eastern oyster, *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers," *BMC Genomics*, vol. 8, article 157, 2007.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

