CHAPTER III

RESULTS

3.1 Haemocytes and total RNA preparation

Haemolymph was collected from 3 month-old *P. monodon* and approximately 1 ml was obtained per individual. The haemolymph was centrifuged to separate haemocytes from plasma and haemocytes were used to prepare total RNA using Trizol reagent. The A260/A280 ratio of total RNA which was prepared by this method were 1.5-1.8 indicating accepted quality of total RNA used in this study. The average total RNA obtained from normal shrimps was approximately 19 μg per individual whereas that obtained from infected shrimp was 11 μg per individual. Therefore, total RNA from 15 healthy shrimps were pooled for used in a construction of the normal library whereas that of 20 infected shrimps were used to establish the *Vibrio*-infected cDNA library.

An ethidium bromide stained 1% agarose-formaldehyde gel of total RNA of normal shrimps revealed two predominant bands; 28S rRNA (4.7 kb) and 18S rRNA (1.9 kb) along with smeared RNA with molecular sizes up to approximately 10 kb (Figure 3.1).

3.2 Construction of haemocyte cDNA libraries

3.2.1 The normal library

The haemocyte cDNA library of normal shrimps was constructed by a non-directional cloning approach using Lambda ZapII cloning system. The cDNAs were cloned into the *Eco*RI-site of λ ZapII. The titering assay of this library showed that its efficiency was approximately 4 x 10^6 pfu. The titer of the library was inceased to approximately 1 x 10^{10} pfu/ml after

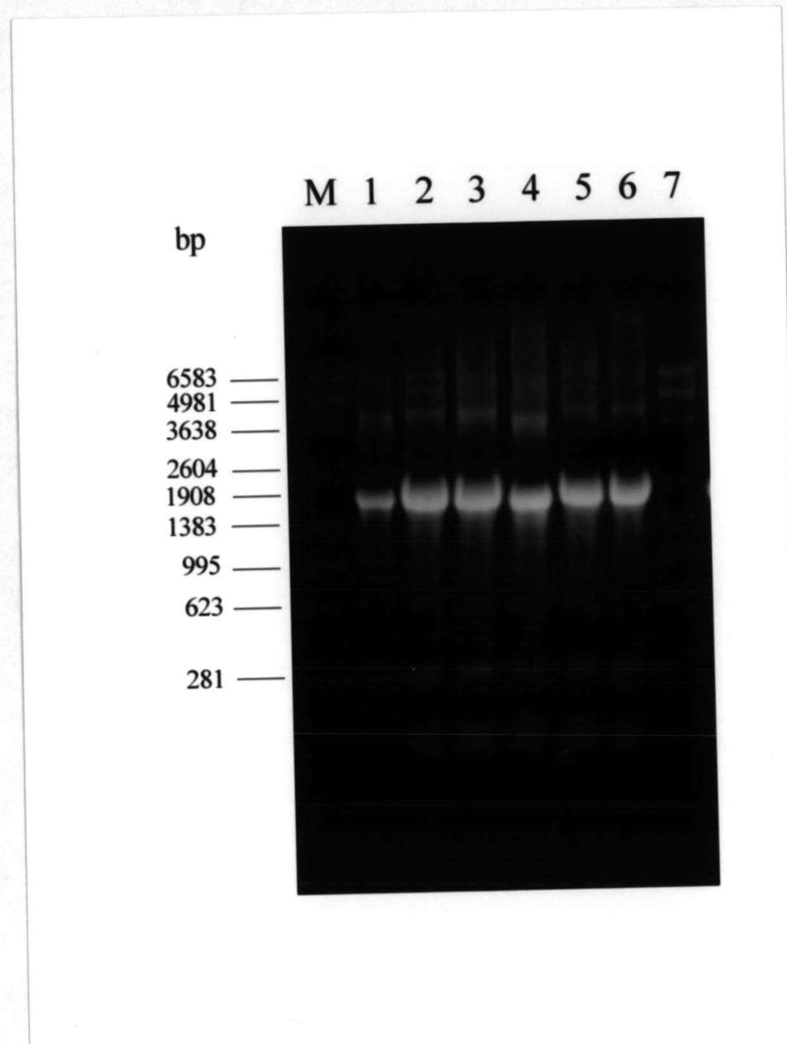

Figure 3.1 Total RNA from haemocytes of 6 normal shrimps (lanes 1-6) electrophoresed on 1% formaldehyde agarose gel. Lane M: RNA marker

amplification. The average insert length of EST clones in this library was 858 bp as determined by colony PCR (Figure 3.2) or enzymatic digestion.

3.2.2 The infected library

The haemocyte cDNA library of infected shrimps was constructed by a directional cloning procedure using a Lambda ZipLox cloning system. The cDNAs, with *Not*I at the 3' end and *Sal*I at the 5' end, were cloned into the *Not*I and *Sal*I arms of the λ ZipLox. The titer of the infected library was 2.5 x 10^5 pfu and become 2 x 10^{10} pfu/ml after amplification. The average insert length of the EST clones in this library was 964 bp as determined by colony PCR or enzymatic digestion (Figure 3.3).

3.3 EST analysis

3.3.1 Homology search

For the normal library, 615 EST clones were randomly selected and partially sequenced at the 5' region of the clones using an automated DNA sequencer (LC4000, LICOR). An average nucleotide sequences of cDNA insert is 607 bp, corresponding to a total insert sequence length of 352 kb. The nucleotide sequences of each clone were analyzed by homology searches against data in the GenBank based on nucleotide similarity (BLASTN) and similarity of translated protein sequence (BLASTX) (Altschul et al., 1990). A putative function of each EST was assigned according to the greatest significant similarity of a particular EST with that of the matched gene in the GenBank.

A total of 317 EST clones (51.5%) represented gene homologues whereas the remaining clones (48.5%) did not match to any sequence in the GenBank. One hundred and nine clones (35%) of the matched EST were identified as known genes by both BLASTN and BLASTX programs while the remaining 65% (205 clones) were recognized only by the BLASTX program (Table 3.1). From the matched EST clones, 231 clones (75%) of which had the orientation of the original mRNA when sequencing with M_{13} forward primer.

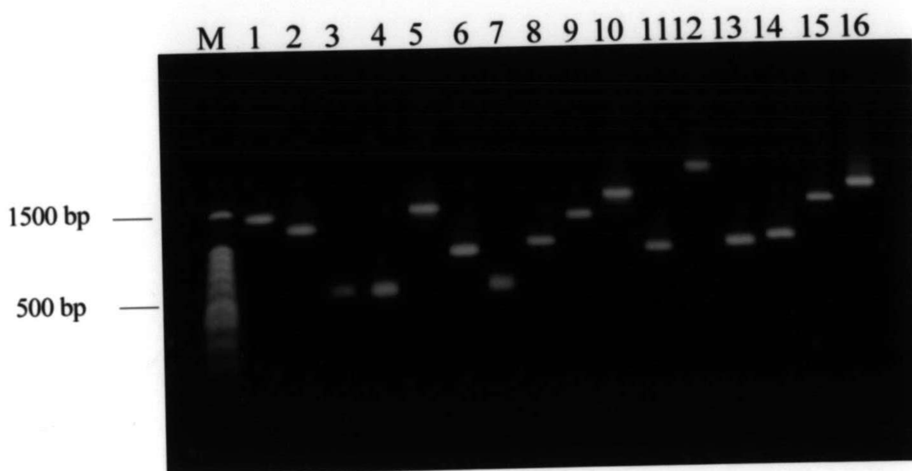

Figure 3.2 Determination of insert sizes of recombinant clones from the normal library using colony PCR. The PCR product form each clone was electrophoresed on 1% agarose gel(lanes 1-16). A DNA ladder 100 bp(M) was used as a DNA marker.

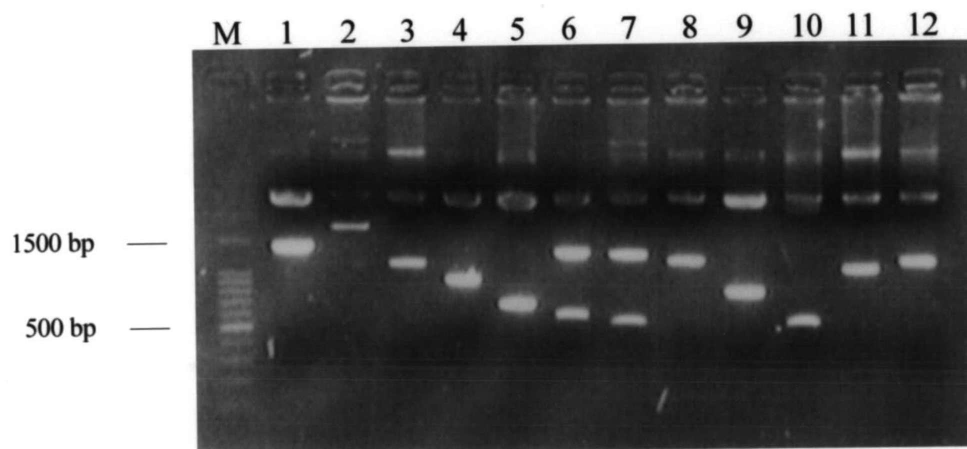

Figure 3.3 Determination of insert sizes of recombinant clones from the *V. harveyi* -infected library by digestion of recombinant plasmids with restriction endonuclease. The digested DNA form each clone was electrophoresed on 1% agarose gel (lanes 1-12). A DNA ladder 100 bp(M) was used as a DNA marker.

For the infected library, 447 EST clones were randomly selected and partial sequenced at the 5' region of the clones. An average nucleotide sequence of cDNA inserts is 549 bp, corresponding to a total insert of 246 kb. The nucleotide sequence of each clone from the infected library was analyzed as described for the normal library.

A total of 215 EST clones (48.1%) showed significant similarity with deposited sequences whereas 232 EST clones (51.9%) did not match with any gene in the GenBank. Homologies of 81clones (38 %) from the matched ESTs were found by both BLASTN and BLASTX while the remaining 62 % (134 clones) were identified only by BLASTX program (Table 3.1).

ESTs in both libraries significantly matched to genes previously identified in many organisms including *Penaeid specises* (49 clones of the normal library and 30 clones of the infected library); *P. monodon* (31 clones in the normal library and 20 clones in the infected library), *P. vannamei* (10 clones in the normal library and 9 clones in the infected libraryt), *P. notialis* (8 clones in the normal library), *P. setiferus* (1 clone in the infected library). The ESTs, which are homologous to genes in other crustaceans, were 41 clones (13.3%) and 31 clones (14.4%) from normal and infected libraries, respectively. Sixty-six clones (21.4%) of normal library and 47 clones (21.9%) of infected library were homologues of genes previously reported in other arthropods. Of these, homologues of *Drosophilia melanogaster* genes were the largest group (38 clones in normal library and 30 clones in infected library).

Homologous genes in other animals were found for 147 clones (47.7%) and 99 clones (46.0%) from normal and infected library, respectively. Within this group, homologous genes in mammalis were the largest group (81 clones in normal library and 64 clones in the infected library). Homologues of genes in plant, fungi and bacteria were found for 5 clones (1.6%) and 8 clones (37.1%) from normal and infected libraries, respectively.

Table 3.1 Summary of sequences and clones represented in the cDNA libraries.

	Normal library	Infected library
No. of clone sequenced	615	447
Matched clones	308(50.1%)	215(48.1)
Penaeid shrimps	49(15.9%)	30(14.0%)
Other crustaceans	41(13.3%)	31(14.4%)
Other arthropods	66(21.4%)	47(21.9%)
Other animals	147(47.7%)	99(46.0%)
Non-animal	5(1.6%)	8(3.7%)
Total nucleotide sequenced (kb)	352	246
Average sequenced length (bp)	607	549

A plot between the number of clones sequenced versus the number of newly identified unique sequences illustrated that newly unique genes can still be further identified because the curves still did not reach a plateau of identification. The ability to isolate new sequences of the normal library was 32%, 40%, 31%, 34%, 25% and 19% when 100, 200, 300, 400, 500 and 600 recombinant clones were sequenced. Likewise, 39%, 23%, 33%, and 17% newly unique sequences of the *Vibrio*-infected library could be identified when 100, 200, 300 and 400 recombinant clones were sequenced. The data suggested a more rapid saturation of the infected library than that of the normal library (Figure 3.4).

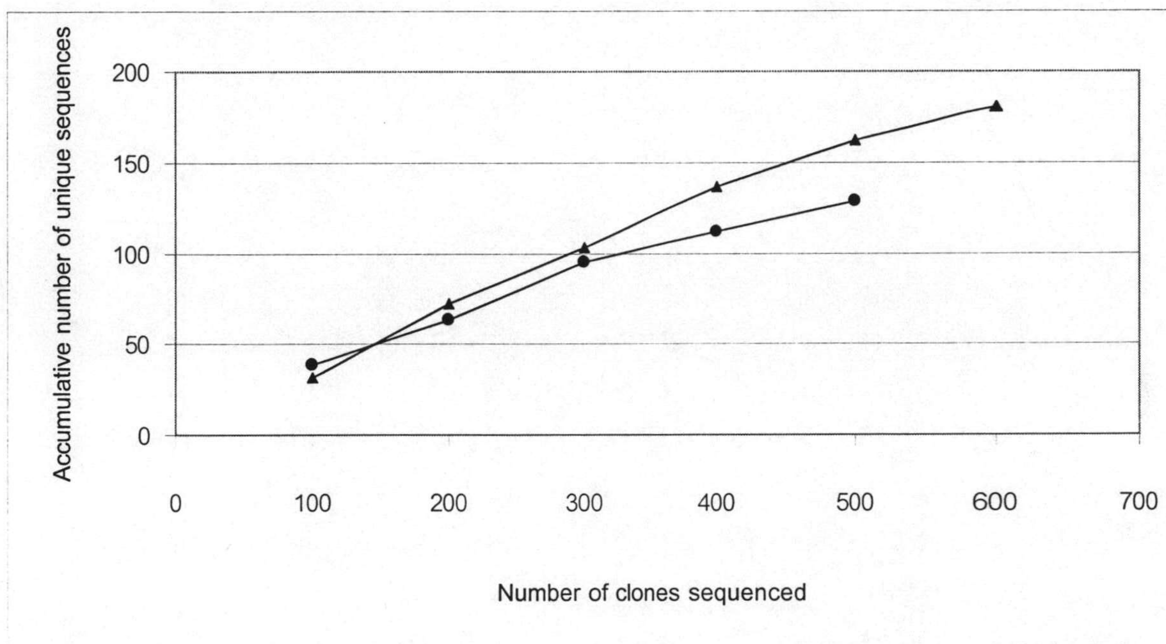

Figure 3.4 The possibility to isolate newly unique sequences of the normal cDNA library (▲) and V. hraveyi-infected library (●) was determined by relationship between numbers of clones sequenced and accumulative numbers of unique sequences obtained.

3.3.2 Classification of putative identified clones

Matched ESTs from both libraries detected as homologues of genes whose the function was well defined were classified into 6 broad functional categories based on significant sequence homology according to the criteria proposed by Adams et al., (1991). These were genes involving with (1) gene expression, regulation and protein synthesis, (2) internal/external structure and motility, (3) metabolism, (4) defense and homeostasis, (5) signaling and communication, and (6) cell division/DNA synthesis, repair and replication. The ESTs, homologue of genes with unknown gene functions were classified unknown genes with the complete open reading frame (ORF) and summarized in Table 3.2 (Details of each clone are shown in appendix A).

Table 3.2 Matched ESTs in each functional category

Functional category	Normal library		Infected library	
	No. in the category	% of ESTs analysed	No. in the category	% of ESTs analysed
1. Gene expression, regulation and protein synthesis	80	13	91	20.4
2. Internal/external structure and motility	47	7.6	11	2.5
3. Metabolism	68	11.1	36	8.1
4. Defense and homeostasis	65	10.6	50	11.2
5. Signaling and communication	12	2.0	3	0.7
6. Cell division/DNA synthesis, repair and replication	6	1.0	5	1.1
Unidentified function	30	4.9	19	4.3
Unknown gene	307	49.9	232	51.9
Unknown with ORF	47	7.6	58	13.0
Total ESTs	615	100	447	100

The percentage of EST clones in each category illustrated that ESTs for gene expression, regulation and protein synthesis were the largest group in both normal and infected libraries. ESTs having cell division/DNA synthesis and signaling/communication function were the smallest group for normal and infected libraries, respectively. Three categorized ESTs: Internal/external structure and motility, Signaling and communication, and unidentified function of the normal library were more abundant than those in the infected library. Conversely, the percentage of EST clones in gene expression, regulation and protein synthesis was more abundantly expressed in the infected library. The percentage of EST clones homologues of genes in metabolism, defense and homeostasis, and cell devision/DNA synthesis, repair and replication from the normal library were comparable to those from the infected library.

Five hundred and thirty-two matched ESTs represent 288 different putative proteins and the frequency of each identified protein in the normal library and the *V. harveyi* infected library are shown in Tables 3.3-3.9.

Gene expression, regulation and protein synthesis

Putative genes of this group are the highest abundant EST clones found in both cDNA libraries. A total of 171 clones (32.7% of the matched EST clones) representing 94 different genes were identified (Table 3.3). Major putative proteins included a group of ribosomal proteins of small (54 clones) and large subunits (61 clones), followed by the group of elongation factors (13 clones), and eukaryotic translation initiation factors (5 clones), respectively. Other genes were ubiquitin-conjugating enzymes, transcription initiation factor, snRNP-associated protein, etc. Redundant clones were found. Among overall sequenced clones, elongation factor-1 alpha showed the highest redundancy (9 clones). This putative gene was the most abundant clone in the normal library whereas 40S ribosomal protein S15 and 60S ribosomal protein L8 were the most abundant clones in the infected library.

Table 3.3 Gene expression and protein coding gene transcripts isolated from normal and *V. harveyi* infected *P. monodon* haemocytes

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
5S ribosomal protein	*Mus musculus*	2.00E-93	98	689	2	-	2
16S ribosomal RNA gene, mitochondria RNA	*Penaeus monodon*	0	99	699	3	-	3
40S ribosomal protein S3	*Oryziaa latipes*	1.00E-93	84	774	4	1	5
40S ribosomal protein S4	*Gallus gallus*	1.00E-105	83	812	2	1	3
40S ribosomal protein S5	*Homo sapiens*	4.00E-74	95	601	-	2	2
40S ribosome protein S7	*Xanopus laevis*	9.00E-75	84	631	1	2	3
40S ribosomal protein S8	*Apis mellifera*	6.00E-78	77	647	3	1	4
40S ribosomal protein S10	*Rattus rattus*	8.00E-50	80	540	-	1	1
40S ribosomal protein S11	*Xenopus laevis*	3.00E-62	85	553	-	2	2
40S ribosomal protein S12	*Homo sapiens*	5.00E-53	85	557	-	4	4
40S ribosomal protein S15	*Homo sapiens*	2.00E-47	68	470	-	5	5
40S ribosomal protein S15A	*Drosophila melanogaster*	2.00E-62	96	409	1	-	1
40S ribosomal protein S16	*Homo sapiens*	4.00E-69	96	620	1	1	2
40S ribosomal protein S17	*Homo sapiens*	3.00E-59	94	450	1	2	3
40S ribosomal protein S18	*Cherax destructor*	1.00E-67	99	504	-	2	2
40s ribosomal protein S18	*Homo saprin*	1.00E-63	94	494	1	-	1
40S ribosomal protein S20	*Xenopus laevis*	8.00E-47	95	455	-	2	2
40S ribosomal protein S24	*Fugu rubripes*	3.00E-50	86	458	2	1	3
40S ribosomal protein S25	*Drosophila melanogaster*	4.00E-29	89	394	-	1	1
40S ribosomal protein S26	*Homo sapein*	8.00E-45	90	579	1	-	1
40S ribosomal protein S27a	*Homo sapein*	1.00E-54	75	465	2	-	2
40S ribosomal protein S28	*Cricetulus griseus*	3.00E-19	97	302	-	1	1
40S ribosomal protein S29	*Homo sapein*	7.00E-12	75	210	1	-	1
40S ribosomal protein S30	*Rattus rattus*	5.00E-24	62	751	3	-	3
40S ribosome protein S32	*Drosophila melanogaster*	2.00E-55	90	473	1	-	1
60S ribosomal protein Po	*Drosophila melanogaster*	3.00E-33	83	441	-	1	1
60S ribosomal protein like	*Arabidopsis thaliana*	1.00E-34	72	474	1	-	1
60S ribosomal protein L3	*Mus musculus*	E-119	86	804	-	2	2

Table 3.3 Gene expression and protein coding gene transcripts isolated from normal and *V. harveyi* infected *P. monodon* haemocytes (continued)

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
60S ribosomal protein L5	*Bombyx mori*	4.00E-59	78	625	1	-	1
60S ribosomal protein L5A	*Xenopus laevis*	E-105	79	600	-	3	3
60A ribosomal protein L6	*Homo sapiens*	6.00E-26	67	707	1	-	1
60A ribosomal protein L7A	*Gallus gallus*	1.00E-62	68	864	1	-	1
60A ribosomal protein L8	*Aedes albopictus*	E-112	94	807	-	5	5
60S ribosomal protein L10	*Drosophila melanogaster*	7.00E-96	95	571	1	1	2
60S ribosomal protein L11	*Drosophila melanogaster*	2.00E-58	92	470	1	-	1
60S ribosomal protein L12	*Rattus rattus*	7.00E-69	89	691	-	1	1
60S ribosomal protein L13	*Gallus gallus*	1.00E-43	72	465	-	2	2
60S ribosomal protein L14	*Rattus norvegicus*	2.00E-29	61	644	1	-	1
ribosomal protein L14	*Xenopus laevis*	2.00E-18	79	312	-	2	2
60S ribosomal protein L17	*Rattus rattus*	2.00E-58	79	620	3	-	3
60S ribosomal protein L17A	*Drosophila melanogaster*	2.00E-68	93	601	-	2	2
60S ribosomal protein L18a	*Homo sapiens*	8.00E-64	80	569	-	3	3
60S ribosomal protein L19	*Drosophila melanogaster*	6.00E-50	74	560	2	-	2
60S ribosomal protein L22	*Tripneuster gratilla*	8.00E-21	60	633	-	1	1
60S ribosomal protein L24	*Homo sapiens*	1.00E-33	70	518	-	3	3
60S ribosomal protein L26	*Homo sapiens*	8.00E-51	83	460	1	-	1
60S ribosomal protein L27A	*Rattus rattus*	1.00E-49	75	516	-	2	2
60S ribosomal protein L28	*Homo sapiens*	1.00E-13	64	362	-	1	1
60S ribosomal protein L29	*spodoptera frugiperda*	6.00E-12	65	330	0	2	2
60S ribosomal protein L30	*Homo sapiens*	1.00E-47	90	411	1	1	2
60S ribosomal protein L31	*Heliothis virescene*	3.00E-32	67	395	-	1	1
60S ribosomal protein L35a	*Homo sapiens*	2.00E-38	79	381	-	2	2
60S ribosomal protein L37	*Homo sapiens*	4.00E-24	74	335	1	-	1
60S ribosomal protein L37A	*Ostertagia ostertagi*	1.00E-28	71	324	1	2	3
60S ribosomal protein L39	*Drosophila melanogaster*	3.00E-19	87	269	-	2	2
60S acidic ribosomal protein P2	Sus scrofa	1.00E-14	79	394	-	2	2

Table 3.3 Gene expression and protein coding gene transcripts isolated from normal and *V. harveyi* infected *P. monodon* haemocytes (continued)

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
60S ribosomal protein CEP52	*Drosophila melanogaster*	1.00E-65	98	520	1	-	1
60S ribosomal protein eL12'	*Artemia sp.*	2.00E-20	89	406	-	2	2
60S ribosomal protein, Large P2	*Homo sapiens*	5.00E-06	65	403	-	1	1
33kDa transcription co-activator	*Homo sapiens*	1.00E-54	78	845	1	-	1
Bmsqd-2	*Bombyx mori*	1.00E-33	89	600	-	1	1
cellular nucleic acid binding protein	*Xenopus laevis*	1.00E-36	52	751	1	-	1
chromatin-specific transcription elongation factor	*Homo sapiens*	9.00E-51	69	827	1	-	1
double stranded RNA binding nuclear protein, ILF3	*Homo sapiens*	2.00E-06	58	535	1	-	1
Ef2b gene product [alt 2]	*Drosophila melanogaster*	2.00E-60	88	673	-	1	1
elongation factor-1 alpha	*Salmo salar*	E-111	90	769	5	4	9
elongation factor-1 beta	*Bombyx mori*	5.00E-37	85	458	1	-	1
elongation factor 2	*Arnadillium vulgare*	E-133	95	820	1	1	2
eukaryotic translation initiation factor XeIF-4A III	*Xenopus laevis*	1.00E-133	97	815	2	-	2
eukaryotic translation initiation factor2, beta subunit2	*Drosophila melanogaster*	5.00E-09	77	520	-	1	1
eukaryotic translation initiation factor3, subunit2	*Mus musculus*	1.00E-39	88	704	1	-	1
eukaryotic translation initiation factor3, subunit5	*Homo sapiens*	4.00E-08	75	296	-	1	1
LMPX of lamprey	*Petromyzon marinus*	2.00E-71	86	718	1	-	1
large subunit ribosomal protein rpl 44	*Ades triseriatus*	1.00E-42	81	369	2	1	3
multicatalytic endopeptidase complex	*Homo sapiens*	2.00E-82	87	760	1	-	1
nascent polypeptide-associated complex alpha chain	*Mus musculus*	1.00E-17	52	619	1	-	1
polyubiquitin	*Cricetulus griseus*	7.00E-47	95	315	1	-	1
probable reverse transcriptase	*Ades triseriatus*	3.00E-05	46	454	1	-	1
Probable threonyl-tRNA synthetase	*Caenorhabditis elegans*	1.00E-112	79	826	1	-	1
protein arginine methy transferase	*Mus musculus*	3.00E-40	52	689	1	-	1
protein synthesis initiation factor 4A	*Mus musculus*	2.00E-84	84	657	1	-	1
putative reverse transcriptase	*Takifugu rubripes*	8.00E-21	56	620	1	-	1
QM protein	*Bombyx mandarina*	1.00E-110	91	690	1	1	2
RpL 19 gene product	*Drosophila melanogaster*	1.00E-63	79	620	-	1	1

Table 3.3 Gene expression and protein coding gene transcripts isolated from normal and *V. harveyi* infected *P. monodon* haemocytes (continued)

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
RpS 9 gene product	*Drosophila melanogaster*	9.00E-78	87	626	1	1	2
RpS 25 gene product	*Drosophila melanogaster*	8.00E-29	89	960	2	-	2
snRNP-associated protein	*Dario rerio*	2.00E-31	74	418	1	-	1
transcription co-repressor Sin3	*Xenopus laevis*	6.00E-11	81	419	-	1	1
transcription initiation factor TFIID 110 kDa subunit	*Drosophila melanogaster*	1.00E-21	54	955	1	-	1
translationally controlled tumor protein homolog	*Drosophila melanogaster*	4.00E-35	72	557	-	2	2
ubiquitin and ribosomal protein S27a precursor	*Homo sapiens*	3.00E-58	77	792	-	1	1
ubiquitin-conjugating enzyme E2-24 kDa	*Drosophila melanogaster*	3.00E-73	99	862	2	-	2
ubiquitin-conjugating enzyme E2L	*Homo sapiens*	4.00E-71	94	792	-	1	1
von Hippel-Lindau binding protein 1	*Homo sapiens*	3.00E-12	84	378	1	-	1

Internal/external structure and motility

A total of 58 clones (11.1 % of the matched EST clones), representing 29 different genes were homologues of genes of internal/external structure and motility, (Table 3.4). The major putative protein found in this category was actin-related proteins (29 clones), followed by a group of tubulin genes (11 clones). Other genes were ubiquitin-like protein SMT2, myosin regulatory light chain, cytoplasmid beta chain, etc. Beta actin was the highest dominant clones (7 clones) in the normal library whereas no predominant clone was found in the infected library.

Metabolism

One hundred and four clones (19.9% of the matched EST clones) representing 55 different putative genes involved in metabolism (Table 3.5). This group includes putative mitochondrial gene products, for example, cytochrome b, cytochrom c subunits, ATP synthease, ATPase subunit b, NADH dehydrogenase subunits, etc. The major transcripts were those of cytochrome C oxidase subunits (35 clones), followed those of ATPase subunits (14 clones) and NADH dehydrogenase subunits (6 clones), respectively. The remaining putative genes identified were aldehyde reductase, steroid dehydrogenase, proteasome subunits, glucose-6 phosphatase, arginase, etc. Cytochrome c oxidase subunit I was the most abundant putative gene in both libraries (13 clones in the normal and 7 clones in the infected library).

Defense and homeostasis

One hundred and fifteen clones (65 clones from the normal library and 50 clones from the infected library) representing 10.8% of the total clones sequenced, are putative immune-related genes. These clones represented 34 different immune genes and were categorized in 5 different subgroups (Table 3.6).

The first subgroup was antimicrobial protein/peptides which were predominated in both libraries. Gene transcripts in this group were composed

Table 3.4 Internal/External structure and motility related gene products isolated from normal and *V. harveyi* infected *P. monodon* haemocytes

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
actin related protein	*Lumbricus rubellus*	7.00E-23	99	677	1	-	1
actin related protein 2/3 complex, subunit3	*Homo sapiens*	1.00E-63	80	829	-	1	1
actin, alpha cardiac muscle	*African clawed frog*	6.00E-67	87	800	1	-	1
actin, clone 403	*Artemia sp*	1.00E-120	99	784	6	-	6
actin, cytoplasmic A3	*Bombyx mori*	1.00E-110	98	642	2	-	2
actin1	*Atlantic horseshoe crab*	1.00E-125	98	800	6	1	7
actin2	*Daucus carota*	1.00E-58	96	731	1	-	1
actin3	*Limulus polyphemus*	1.00E-109	98	671	1	-	1
alpha-2-tubulin	*Gecarcinus lateralis*	1.00E-29	66	542	2	-	2
beta actin	*Penaeus vannamei*	1.00E-156	96	909	7	1	8
beta tub 56D gene product (alt1)	*Drosophila melanogaster*	1.00E-141	97	938	1	-	1
calponin homolog	*Schistosoma japonicus*	2.00E-27	63	615	1	-	1
cytoplasmic beta chain	*Xenopus laevis*	1.00E-126	95	907	2	-	2
dynein light chain, cytoplasmic	*Drosophila melanogaster*	3.00E-45	99	729	1	-	1
feritin 2 light chain homolog	*Drosophila melanogaster*	2.00E-04	48	904	1	-	1
gamma-actin	*Mus musculus*	1.00E-44	100	402	-	1	1
INNEXIN INX3	*Drosophila melanogaster*	1.00E-34	55	0	-	1	1
myosin regulatory light chain	*Drosophila melanogaster*	1.00E-69	95	536	1	2	3
Plelota protein	*Drosophila melanogaster*	2.00E-54	73	558	-	1	1
profilin	*Drosophila melanogaster*	3.00E-20	59	800	2	-	2
SMC 1 protein	*Drosophila melanogaster*	1.00E-85	76	930	1	-	1
stromelysin-3 precusor	*Xenopus laevis*	3.00E-25	49	661	1	-	1
tubulin alpha-1- chain	*Homarus americanus*	1.00E-171	92	985	2	2	4
tubulin alpha-3 chain	*Homarus americanus*	3.00E-76	91	522	1	-	1
tubulin beta	*Homo sapiens*	1.00E-107	72	885	3	-	3
Ubiquitin-like protein SMT2	*Caenorhabditis elegans*	4.00E-31	83	849	1	-	1
ras-related protein RAB-1A	*Lymnaea stagnalis*	8.00E-96	90	923	1	-	1
SEC 61, gamma subunit	*Mus musculus*	2.00E-10	59	353	-	1	1
protein complex subunit34	*Homo sapiens*	2.00E-19	84	738	1	-	1

Table 3.5 Metabolism-related transcripts isolated from normal and *V. harveyi* infected *P. monodon*

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
ATP syntase coupling factor 6,mitochondrial precu	*Drosophila melanogaster*	1.00E-22	73	432	-	1	1
ATP synthase alpha chain, mitochondrial precursoı	*Drosophila melanogaster*	2.00E-14	80	630	1	-	1
ATP synthase FO subunit 6	*Penaeus monodon*	3.00E-85	91	823	5	2	7
ATP synthase oligomycin sensitivity conferral protein precursor	*Drosophila melanogaster*	8.00E-57	75	758	-	3	3
ATP synthase subunit C	*Manduca sexta*	4.00E-12	61	777	2	-	2
ATPase subunit 6	*Penaeus notialis*	3.00E-72	70	748	2	-	2
ATPsyn-gamma gene product	*Drosophila melanogaster*	3.00E-69	71	710	1	-	1
cytochrome B	*Drosophila melanogaster*	E-105	74	867	2	2	4
cytochrome C oxidase subunit VI precusor	*Thunnus niloticus*	2.00E-18	76	503	1	-	1
cytochrome C oxidase subunit I	*Penaeus monodon*	E-135	93	809	13	7	20
cytochrome C oxidase subunit II	*Penaeus monodon*	8.00E-98	85	671	3	3	6
cytochrome C oxidase subunit III	*Penaeus monodon*	1.00E-105	84	765	6	-	6
cytochrome C oxidase subunit VIb	*Saccharomyces cerevisiae*	1.00E-20	70	552	-	1	1
cytochrome C oxidase subunit VIIc	*Rattus norvegicus*	5.00E-09	72	417	1	-	1
NADH dehydrogenase subunit I	*Penaeus monodon*	E-104	94	676	-	1	1
NADH dehydrogenase subunit II	*Penaeus monodon*	1.00E-100	77	804	1	-	1
NADH dehydrogenase subunit IV	*Penaeus monodon*	6.00E-80	78	611	1	1	2
NADH dehydrogenase subunit V	*Penaeus monodon*	1.00E-108	88	898	1	-	1
NADH dehydrogenase subunit VI	*Penaeus monodon*	3.00E-32	75	385	-	1	1
NADH-Ubiquinone oxidoreductase 42 kDa subunit mitochondrial	*Drosophila melanogaster*	6.00E-37	75	710	2	-	2
2,4-dienoyl-CoA reductase (NADH)	*Rattus norvegicus*	7.00E-93	73	507	-	1	1
acyl coenzyme A dehydrogenase, long chain	*Rattus norvegicus*	1.00E-81	86	930	1	-	1
ADP-ribosylation factor6	*Gallus gallus*	3.00E-97	97	661	-	1	1
aldehyde reductase	*Sus scrofa*	6.00E-37	66	708	1	-	1
amino acid starvation-induced protein	*Rattus norvegicus*	2.00E-33	68	534	-	1	1
Apg 12	*Homo sapiens*	2.00E-32	72	724	1	-	1
arginase	*Xenopus laevis*	2.00E-45	65	638	-	1	1
arginine kinase	*Callinectes sapidus*	3.00E-38	98	527	-	1	1

Table 3.5 Metabolism-related transcripts isolated from normal and *V. harveyi* infected *P. monodon* (continued)

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
calmodulin	*Drosophila melanogaster*	2.00E-80	100	747	2	-	2
chain A, Triosephosphate isomerase	*Gallus gallus*	3.00E-54	70	654	1	-	1
chain H, cytochrome Bc1 complex	*Gallus gallus*	4.00E-14	66	593	1	-	1
glucose-6phophatase	*Mus musculus*	2.00E-33	58	768	2	-	2
glutamyl-tRNA (Glu) aminotransferase	*Deinococcus radiodurans*	5.00E-10	56	557	-	1	1
glutathione s-transferase	*Rattus norvegicus*	8.00E-04	54	370	-	1	1
GTP-binding nuclear protein RHEB homolog	*Drosophila melanogaster*	4.00E-70	83	726	2	-	2
guanine nucleotide-binding protein beta subunit	*Oreochromis niloticus*	3.00E-82	88	747	2	-	2
guanine nucleotide-binding protein gamma-1	*Drosophila melanogaster*	3.00E-14	77	397	-	1	1
Na/K-ATPase beta subunit isoform 3	*Drosophila melanogaster*	8.00E-24	55	692	-	1	1
non-selenium glutathione phospholipid hydroperoxidase	*Sus scrofa*	1.00E-29	57	572	1	-	1
nucleoside diphosphate kinase	*Columba livia*	4.00E-49	84	456	1	-	1
peroxisomal Ca-dependent solute carrier	*Orycotolagus cuniculus*	4.00E-63	75	942	1	-	1
PKCq-interacting protein PICOT	*Homo sapiens*	3.00E-52	67	707	1	-	1
protein-glutamine	*Oryctolaqus cuniculus*	3.00E-08	53	750	1	-	1
pterin-4a-carbinolamine dehydratase	*Drosophila melanogaster*	1.00E-18	76	280	1	-	1
putative steroid dehydrogenase	*Mus musculus*	8.00E-35	64	728	1	-	1
RIKEN cDNA 2510049/19 gene	*Mus musculus*	5.00E-18	60	432	-	1	1
stromal cell derived factor2	*Mus musculus*	2.00E-39	59	800	1	-	1
unspecific monooxygenase	*Nicotiana tabacum*	9.00E-19	60	689	-	2	2
vacuolar ATP synthase subunit G	*Manduca sexta*	2.00E-20	68	626	1	-	1
26S proteasome regulatory subunit	*Anopheles stephens*	5.00E-65	86	760	1	-	1
proteasome 25 kDa subunit	*Drosophila melanogaster*	6.00E-72	79	520	-	1	1
proteasome beta chain precursor	*Mus musculus*	9.00E-21	65	673	1	-	1
proteasome subunit alpha type 6	*Mus musculus*	2.00E-30	80	816	1	-	1
proteasome subunit beta type 2	*Homo sapiens*	3.00E-53	69	764	1	-	1
putative senescence-associated protein	*Pisum sativum*	2.00E-29	84	669	-	1	1

Table 3.6 Immune related genes isolated from normal and *Vibrio harveyi* infected shrimps *P. monodon*

Putative identification	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
1. Antimicrobial molecules							
antilipopolysaccharide factor	*Atlantic horseshoe crab*	7.00E-19	64	512	5	16	21
Crustin	*Litopenaeus setiferus*	1.00E-42	65	515	5	8	13
Penaeidin-2 precursor	*Litopenaeus setiferus*	2.00E-05	57	651	1	3	4
Penaeidin-3c precursor	*Litopenaeus setiferus*	1.00E-18	66	412	6	2	8
Penaedine-3k precursor	*Litopenaeus setiferus*	9.00E-11	54	647	-	1	1
P lysozyme structural	*Mus musculus*	2.00E-26	67	541	2	2	4
2. ProPO systems and oxidative enzyme							
Clottable protein	*Penaeus monodon*	E-130	98	458	3	3	6
Cytosolic manganese superoxidase Dismutase precursor	*Callinectes sapidus*	E-129	86	930	-	1	1
Glutathione peroxidase	*Homo sapiens*	3.00E-18	76	331	1	-	1
Haemocyte protein-glutamine gamma-glutamyltransferase	*Tachypleus tridentatus*	2.00E-48	76	746	4	-	4
peroxidase	*Aedes aegypti*	1.00E-21	52	747	1	-	1
Prophenoloxidase	*Penaeus monodon*	1.00E-51	80	465	1	-	1
prophenoloxidase activating factor	*Holotrichia diomphalia*	7.00E-18	70	550	1	2	3
3. Proteinases and inhibitor							
Gene MAC25 protein (Kazal proteinase inhibitor homology)	*Homo sapiens*	9.00E-10	53	405	-	3	3
Haemocyte protease-1	*Manduca sexta*	1.00E-12	57	700	1	-	1
proteinase inhibitor-signal crayfish	*Pacifastacus leniusculus*	5.00E-31	60	503	8	1	9
Cathepsin B-like cytoeine proteinase, 29 k precursor	*Sarcophaga perregrina*	6.00E-18	88	659	1	-	1
Serine protease	*Pacifastacus leniusculus*	4.00E-79	79		1	1	2
Whey acidic protein (putative Protease inhibitor)	*Trichosurus vulpecula*	5.00E-06	48	712	1	1	2
4. Heat shock protein							
Heat shock protein 10	*Gallus gallus*	2.00E-25	73	665	1	-	1
Heat shock cognate 70 kD protein	*Oncorhynchus mykiss*	9.00E-95	76	772	1	1	2
Heat shock protein 70	*Hydra magripapillata*	2.00E-56	90	806	1	1	2
Heat shock protein 90	*Gallus gallus*	9.00E-27	70	618	2	-	2
5. Other immune molecule							
Fc fragment of IgE	*Homo sapiens*	5.00E-07	67	533	-	1	1
Protein c	*Mouse*	3.00E-10	63	708	1	-	1
Protein-kinase c inhibitor	Bovine	2.00E-38	78	756	1	-	1
Thymosin beta-9 and beta 8	Bovine	2.00E04	77	706	2	-	2
Thymosin beta-11	*Oncorhynchus mykiss*	4.00E-27	64	872	3	-	3
Perlucin	*Haliotis cuniculus*	7.00E-16	54	575	2	-	2
Cyclophilin 18	*Oryctolagus cuniculus*	8.00e-58	72	756	1	-	1
Chaperonin containing t-complex Polypeptide 1	*Homo sapiens*	9.00E-67	77	637	3	-	3
Catalase	*Campylobacter jejuni*	3.00E-09	93	575	1	-	1
Peptide-prolyl cis-trans isomer 5	*Drosophila melanogaster*	3.00E-68	81	816	4	3	7

of anti-lipopolysaccharide factor (ALF), crustins, penaeidins and lysozyme. The second subgroup was those in the Prophenoloxidase systems and oxidative enzymes; Prophenoloxidase (proPO), prophenoloxidase activating factor (ppAA), haemocyte-glutamine gamma-glutamyl transferase (TGase), clottable protein, gluthathione peroxidase, peroxidase, cytosolic manganese superoxide dismutase. The third subgroup of immune-related genes was proteinases and their inhibitors. They were proteinase inhibitors, gene MAC 25 protein (Kazal proteinase inhibitor homologue), whey acidic protein, serine protease, haemocyte protease-1 and probable cathepsinB-like cystein proteinase. The fourth subgroup was heat shock proteins (HSP). Gene transcripts in this group are composed of those for HSP10, HSP70 and HSP90. The last group includes other immune-related molecules such as perluci, thymosin beta-11, thymosin beta-9 and beta-8, and protein C etc.

The predominant clone of these immune genes in the normal library is proteinase inhibitor (8 clones, 12.3%of immune genes in the normal library). Interestingly, the EST homologues of ALF were greatly increased in the infected library (16 clones, 32% of the total immune genes in infected library).

Signaling and communication

Fifteen clones (28% of the match EST clones) representing 15 different genes identified as putative genes of signaling and communication were identified (Table 3.7). Thus, redundancy of recombinant clones was not found in this group. Several receptors such as a sialoglycoprotein receptor, GABA-A receptor were found.

Cell division/DNA synthesis, repair and replication

This group was the smallest group, composed of 11 clones (1.1% of the matched EST clones), representing 4 different genes (Table 3.8). The predominant proteins are histone 1 (5 clones), followed by histone 3 (4 clones).

Table 3.7 Signaling and communication isolated-related transcripts from normal and *V. harveyi* infected *P. monodon* haemocytes

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
11-1 polypeptide	*Plasmodium faciparum*	6.00E-04	56	565	1	-	1
agrin precursor	*Homo sapiens*	6.00E-06	48	428	1	-	1
asialoglycoprotein receptor	*Mus musculus*	2.00E-05	45	922	1	-	1
beta 2-chimeric	*Mus musculus*	2.00E-39	63	799	1	-	1
COP9 (constitutive photomorphogenic) subunit 4	*Mus musculus*	5.00E-78	87	829	1	-	1
gamma-aminobutyric acid (GABA-A) recptor, subunit epsilon	*Rattus norvegicus*	1.00E-07	45	677	-	1	1
kupffer cell receptor	*Rattus norvegicus*	1.00E-13	44	670	1	-	1
LIV-1 protein, estrogen regulated	*Homo sapiens*	1.00E-04	67	934	1	-	1
low-density lipoprotein receptor related proteion-deleted in tumor	*Homo sapiens*	1.00E-12	51	751	1	-	1
moesin / ezrin / radixin	*Drosophila melanogaster*	2.00E-48	72	934	1	-	1
PDGF/VEGF-llike protein	*Drosophila melanogaster*	2.00E-09	52	696	-	1	1
putative VLDL lipoprotein receptor precursor	*Mus musculus*	2.00E-05	44	730	-	1	1
sex-lethal protein homolog	*Ceratitis capitata*	1.00E-50	74	615	1	-	1
shak-8 (lethal) protein	*Drosophila melanogaster*	5.00E-17	65	788	1	-	1
translocon-associated protein	*Canis familiaris*	8.00E-39	73	499	1	-	1

Table 3.8 Cell division/DNA synthesis, repair and replication-related transcripts isolated from normal and *V.harveyi* infected *P. monodon* haemocytes

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
H3 histone	*Homo sapiens*	9.00E-63	96	827	2	2	4
Histone 1	*Mytilus edulis*	3.00E-20	75	542	2	3	5
tankyrase	*Homo sapiens*	8.00E-11	44	436	1	-	1
SPARC-related protein	*Mus musculus*	3.00E-05	46	657	1	-	1

Non-identified functional genes

A total of 49 clones (9.4%) were homologous to genes which non-identified function (Table 3.9). Hypothetical proteins are the largest group. In a total, 12 clones are from the normal library and 6 clones were found from the infected library. Other gene transcripts classified to this group were CG gene products of *D. Melanogaster* KIAA protein etc.

Unknown genes

Unmatched ESTs; 307 clones (49.9%) and 232 clones (51.9%) from the normal and infected libraries, were respectively identified. Using the software Genetyx-Win program to examine the ORF of unmatched ESTs sizes greater than 300 bp in length revealed that 47 ESTs from the normal library and 58 ESTs from the infected had ORF of the genes (Table 3.2).

Table 3.9 Un-identified functional transcripts isolated from normal and *V. harveyi* infected *P. monodon* haemocytes

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
15 kDa selenoprotein	*Homo sapiens*	8.00E-15	65	332	1	-	1
hypothetical 68.8kD protein	Yeast	5.00E-04	65	800	2	-	2
hypothetical 39.6kD protein	Yeast	4.00E-20	60	576	1	-	1
hypothetical 36.9 kDa protein	*Caenorhabditis elegans*	5.00E-10	63	515	1	2	3
KIAA1177 protein	*Homo sapiens*	2.00E-21	78	408	1	-	1
hypothetical protein F 26E4.9	*Caenorhabditis elegans*	3.00E-19	69	564	1	-	1
hypothetical protein 18K protein	*goldfish-mitochondrion*	1.00E-04	65	920	2	-	2
hypothetical protein R17.2	*Caenorhabditis elegans*	3.00E-13	55	636	1	-	1
KIAA0924 protein	*Homo sapiens*	2.00E-07	64	666	1	-	1
KIAA0152 gene product	*Homo sapiens*	1.00E-42	72	531	1	-	1
KIAA0670 protein	*Homo sapiens*	7.00E-25	47	737	1	-	1
selenoprotein w muscle 1	*Rattus norvegicus*	2.00E-23	74	829	3	-	3
hypothetical protein	*Drosophila melanogaster*	3.00E-42	58	872	2	-	2
Ha3611gene product	*Homo sapiens*	2.00E-08	87	371	1	-	1
CG6866 gene product	*Drosophila melanogaster*	7.00E-18	49	710	1	-	1
po 1 polyprotein	*Volvox carteri*	5.00E-04	84	573	1	-	1
tPhLP	*Mus musculus*	7.00E-39	55	834	1	-	1
probeta 2 gene product	*Drosophila melanogaster*	2.00E-62	78	724	1	-	1
CG14429 gene product	*Drosophila melanogaster*	3.00E-18	82	767	1	-	1
Mo23 gene product	*Drosophila melanogaster*	1.00E-20	95	941	1	-	1
hypothetical protein F08F1.8	*Caenorhabditis elegans*	1.00E-12	52	898	1	-	1
hypothetical protein Y45F10D.4	*Caenorhabditis elegans*	7.00E-58	91	875	1	-	1
hrp 65-3 isoform	*Chironomus tentans*	8.00E-06	45	843	1	-	1
brain protein I3	*Homo sapiens*	9.00E-14	71	723	2	-	2
hypothetical 36.9 kDa protein	*Caenorhabditis elegans*	1.00E-13	63	483	-	1	1
hypothetical protein	Homo sapiens	4.00E-04	57	463	-	1	1
KIAA 1594 protein	*Homo sapiens*	1.00E-18	60	372	-	1	1
hypothetical protein FJ12878	*Homo sapiens*	2.00E-58	58	845	-	1	1

Table 3.9 Un-identified functional transcripts isolated from normal and *V. harveyi* infected *P. monodon* haemocytes (continued)

Gene	Closest species	Probability	Matched (%)	Sequence Length (bp)	Frequency (normal shrimp)	Frequency (infected shrimp)	Total clones
unnamed protein product	*Mus musculus*	2.00E-04	51	616	-	1	1
ebiP7015	*Anopheles gambiae*	3.00E-40	84	805	-	1	1
CG14206 gene product	*Drosophila melanogaster*	4.00E-48	79	543	-	1	1
BcDNA GH02921	*Drosophila melanogaster*	2.00E-21	52	557	-	1	1
CG6001 gene product	*Drosophila melanogaster*	1.00E-105	85	922	-	1	1
CG9354 genr product	*Drosophila melanogaster*	5.00E-29	73	386	-	1	1
CG8444 gene product	*Drosophila melanogaster*	2.00E-11	59	636	-	1	1
PTD007 protein	*Homo sapiens*	2.00E-11	51	568	-	1	1
p105 coactivator	*Rattus norvegicus*	5.00E-14	72	764	-	1	1
AK012526 putative	*Mus musculus*	4.00E-06	65	362	-	1	1
AK002940 putative	*Mus musculus*	6.00E-41	78	423	-	1	1
hypothetical protein F26E4.9	*Caenorhabditis elegans*	1.00E-20	70	132	-	1	1
hypothetical protein; CGI-117 protein	*Homo sapiens*	1.00E-08	81	519	-	1	1
hypothetical protein 51.8 kDa protein	*Leishmania major*	1.00E-04	62	586	-	1	1
RE63456p	*Drosophila melanogaster*	7.00E-30	75	367	-	1	1

3.4 Full-length of immune related genes

Nucleotide sequences that were homologous with defense and homeostasis or immune related genes were searched for the open reading frame (ORF) using the Genetyx-Win. 6 different genes contained complete ORF. They were ALFs, penaeidin, proteinase inhibotors, crustins, a heat shock protein 10 and a cytosolic manganese superoxide dismutase.

Anti-lipopolysaccharide factor

An antilipopolysaccharide factor (ALF) is a basic protein initially found in haemocytes of the horseshoe crab. It inhibits the growth of gram-negative bacteria. In this study, 21 clones (5 clones form the normal library and 16 clones from the infected library) were identified as ALF homologues and 17 clones of these contained complete ORFs. Deduced amino acid sequences suggested that at least 4 types (recognized as ALF*Pm*1-ALF*Pm*4) were existent in *P. monodon* haemocytes. They contained ORFs of 252, 360, 369, and 396 bp encoding 84, 120, 123, and 132 amino acids, respectively (Figure 3.5).

Among clones representing full length ALF, 13 (including 5 clones from the normal library) clones were ALF*Pm*3 whereas the remaining clones were classified as each of ALF*Pm*1, ALF*Pm*2 and ALF*Pm*4, respectively. Nucleotide sequences of ALF were aligned. (Figure 3.5). A putative signal peptide peptide at the NH_2-terminus was located between Ala/Gly (A/G) and Gln (Q) (Figure 3.6). Amino acid sequences of *P. monodon* ALF and *T. tridentatus* were 57% - 65% homology (Figure 3.7). Sequence divergence between different family of *P. monodon* ALF were 0.0330 (ALF*Pm*1-ALF*Pm*2) – 1.6257 (ALF*Pm*4-ALF*Pm*2) (Table 3.10). A phylogenetic tree for ALF revealed that gene duplication of ALFPm1 and ALFpm2 was more recent than ALFPm3 and ALFPm 4 which may have dissociated earlier. All ALF*Pm* families were distantly related with *T. tridentatus* ALF (Figure 3.8).

```
ALFPm3          ATGCGTGTGTCCGTGCTGGTAAGCCTGGTGCTGGTGGTGTCCCTGGTGGCACTCTTCGCC
ALFPm4          ATGCGTGTGTCCGTGCTGGTAAGCCTGGTGCTGGTGGTGTCCCTGGTGGCAGTCTTCGCC
ALFPm1          ATGCGAGTCT------TGGTCAGCTTTTTAATGGCACTCAGCCTGATTGCACTTATG---
ALFPm2          ATGCGAGTCT------TGGTCAGCTTTTTAATGGCACTCAGCCTGATTGCACTTATG---
ALFTtr          ------------------------------------------------------------

ALFPm3          CCACAGTGCCAGGCTCAAGGGTGGAGGCTGTGGCAGCGGCCGTCGCCAGCAAGATCGTA
ALFPm4          CCACAGTGCCAGGCTCAAGGGTGGAGGCTGTGGCAGCGGCCGTCGCCAGCAAGATCGTA
ALFPm1          CCACGGTGCCAGGGTCAAGGCGTGCAGGACCTCCTCCCTGCCTTAGTAGAAAAGATCGCT
ALFPm2          CCACGGTGCCAGGGTCAAGGCGTGCAGGACCTCCTCCCTGCCTTAGTAGAAAAGATCGCT
ALFTtr          ------------GATGGTATTTGGACTCAATTGATTTTTACTTTGGTTAATAATTTGGCT
                            * *      *       *        * * *     **  * *

ALFPm3          GGGTTGTGGAGGAACGAAAAAACTGAACTTCTCGG-CCACGAGTGCAAGTTCACCGTCAA
ALFPm4          GGGTTGTGGAGGAACGAAAAAACTGAACTTCTCGG-CCACGAGTGCAAGTTCACCGTCAA
ALFPm1          GGGTTGTGGCACT-CGGATGAGGTGGAGTTCTTGGGCCACAGTTGCAGGTACAGTCAGCG
ALFPm2          GGGTTGTGGCACT-CGGATGAGGTGGAGTTCTTGGGCCACAGTTGCAGGTACAGTCAGCG
ALFTtr          ACTTTGTGGCAAT-CTGGTGATTTTCAATTTTTGGATCATGAATGTCATTATAGAATTAA
                   ******     *     *  *  * ** * **  **    **    *  *

ALFPm3          GCCTTATTTGAAGAGATTCCAGGTGTACTACAAGGGGAGGATGTGGTGCCCAGGCTGG-A
ALFPm4          GCCTTATTTGAAGAGATTCCAGGTGTACTACAAGGGGAGGATGTGGTGCCCAGGCTGGGA
ALFPm1          CCCTTCCTTCTATAGGTGGGAGCTGTACTTCAATGGAAGGATGTGGTGTCCAGGATGG-G
ALFPm2          CCCTTCCTTCTATAGGTGGGAGCTGTACTTCAATGGAAGGATGTGGTGTCCAGGATGG-G
ALFTtr          ACCAACTTTTAGAAGATTGAAATGGAAATATAAAGGTAAATTTTGGTGTCCATCTTGG-A
                 **    **    ** *   *   * * *  ** ** *   * ***** ***    ***

ALFPm3          CGGCCATCAGAGGAGAAGCCAGCACACGCAGTCAGTCCGGGGTAGCTGGAAAGACAGCCA
ALFPm4          CGGCCATCAGAGGAGAAGCCAGCACACGCAGTCAGTCCGGGGTAGCTGGAAAGACAGCCA
ALFPm1          CTCCCTTCACTGGCCGATGTGAGTGA----------------------------------
ALFPm2          CTCCCTTCACTGGCCGATCTCGGACCCGCAGCCCCTCCGGCGCCATAGAGCACGCGACGA
ALFTtr          CTTCTATTACTGGTAGAGCTACTAAATCTTCTAGATCTGGTGCTGTTGAACATTCTGTTA
                *  *  * *  **   *

ALFPm3          AAGACTTCGTTCGGAAAGCTTTCCAGAAAGGTCTCATCTCTCAACAGGAGGCCAACCAGT
ALFPm4          AAGACTTCGTTCGGAAAGCTTTCCAGAAAGGTCTCATCTCTCAACAGGAGGCCAACCAGT
ALFPm1          ------------------------------------------------------------
ALFPm2          GGGACTTCGTGCAGAAGGCGCTGCAGAGTAATCTCATCACGGAGGAAGACGCCAGGATTT
ALFTtr          GAAATTTTGTTGGTCAAGCTAAATCTTCTGGTTTGATTACTCAAAGACAAGCTGAACAAT

ALFPm3          GGCTCAGCTCATAG----------------------------------------------
ALFPm4          GGCTCAGCTCATAGGCCTTTTGCTCTATGAAGAATTGTCAGTGTTCAGCTGCAGTTGGCA
ALFPm1          ------------------------------------------------------------
ALFPm2          GGCTTGAGCACTAA----------------------------------------------
ALFTtr          TTATTTCTCAATATAAT-------------------------------------------

ALFPm3          ------------------------------------------------------------
ALFPm4          ATGGAAGCTCTACCATTTTGATTTCTTGTGTTTTTCCTTTCAATACTGAACCGAAGAGTT
ALFPm1          ------------------------------------------------------------
ALFPm2          ------------------------------------------------------------
ALFTtr          ------------------------------------------------------------

ALFPm3          -------------------
ALFPm4          GAGATATTCATTATGTTAA
ALFPm1          -------------------
ALFPm2          -------------------
ALFTtr          -------------------
```

Figure 3.5 Nucleotide sequence alignment of ALF*Pm*1-ALF*Pm*4 of *P. monodon* with ALF*Ttr* of the horseshoe crab *T. tridentatus*

```
ALFPm4          MRVSVLVSLVLVVSLVALFAPQCQAQGWEAVAAAVASKIVGLWRNEKTELLGHECKFTVK
ALFPm3          MRVSVLVSLVLVVSLVAVFAPQCQAQGWEAVAAAVASKIVGLWRNEKTELLGHECKFTVK
ALFPm2          MR--VLVSFLMALSLIALM-PRCQGQGVQDLLPALVEKIAGLWHSDEVEFLGHSCRYSQR
ALFPm1          MR--VLVSFLMALSLIALM-PRCQGQGVQDLLPALVEKIAGLWHSDEVEFLGHSCRYSQR
                **  ****:::.:**:*:: *:**.** : : .*:..**.***:.::.*:***.*::: :

ALFPm4          PYLKRFQVYYKGRMWCPGWTAIRGEASTRSQSGVAG--KTAKDFVRKAFQK---------
ALFPm3          PYLKRFQVYYKGRMWCPGWDGHQ-RRSQHTQS-VRGSWKDSQRLRSESFPERSHLSTGGQ
ALFPm2          PSFYRWELYFNGRMWCPGWAPFTGRSRTRSPSGAIE--HATRDFVQKALQS---------
ALFPm1          PSFYRWELYFNGRMWCPGWAPFTGRCEX--------------------------------
                * : *:::*::********       .

ALFPm4          ------GLISQQEAN------QWLSS-----------
ALFPm3          PVAQLIGLLLYEELSVFSCSWQWKLYHFDFLCFSFQY
ALFPm2          ------NLITEEDAR------IWLEH-----------
ALFPm1          -------------------------------------
```

Figure 3.6 Nucleotide sequence alignment of ALF*Pm*1-ALF*Pm*4 of *P. monodon*. Putative signal peptide is underlined.

```
ALFPm3          MRVSVLVSLVLVVSLVALFAPQCQAQGWEAVAAAVASKIVGLWRNEKTELLGHECKFTVK
ALFPm4          MRVSVLVSLVLVVSLVAVFAPQCQAQGWEAVAAAVASKIVGLWRNEKTELLGHECKFTVK
ALFPm1          MR--VLVSFLMALSLIALM-PRCQGQGVQDLLPALVEKIAGLWHSDEVEFLGHSCRYSQR
ALFPm2          MR--VLVSFLMALSLIALM-PRCQGQGVQDLLPALVEKIAGLWHSDEVEFLGHSCRYSQR
ALFTtr          ------------------------DGIWTQLIFTLVNNLATLWQSGDFQFLDHECHYRIK
                                              :  ::..::. **:. . ::*.*.*::  :

ALFPm3          PYLKRFQVYYKGRMWCPGWTAIRGEASTRSQSGVAG--KTAKDFVRKAFQK---------
ALFPm4          PYLKRFQVYYKGRMWCPGWDGHQ-RRSQHTQS-VRGSWKDSQRLRSESFPERSHLSTGGQ
ALFPm1          PSFYRWELYFNGRMWCPGWAPFTGRCE---------------------------------
ALFPm2          PSFYRWELYFNGRMWCPGWAPFTGRSRTRSPSGAIE--HATRDFVQKALQS---------
ALFTtr          PTFRRLKWKYKGKFWCPSWTSITGRATKSSRSGAVE--HSVRNFVGQAKSS---------
                * : * :  ::*::***.*      .

ALFPm3          ------GLISQQEANQWLSS--------------------------
ALFPm4          PVAQLIGLLLYEELSVFSCSWQWKLYHFDFLCFSFQYTEELRYSLC
ALFPm1          ----------------------------------------------
ALFPm2          ------NLITEEDARIWLEH--------------------------
ALFTtr          ------GLITQRQAEQFISQYN------------------------
```

Figure 3.7 Amino acid sequence alignment of ALF*Pm*1-ALF*Pm*4 of *P. monodon* with ALFT*tr* of the horseshoe crab *T. tridentatus*

Table 3.10 Genetic distance calculated from amino sequence divergence of ALF in *P. monodon* and an outgroup (ALF*Ttr*)

	Genetic distance			
ALF*Pm*3	ALF*Pm*3	ALF*Pm*4	ALF*Pm*1	ALF*Pm*2
ALF*Pm*4	0.5409			
ALF*Pm*1	1.1026	1.1648		
ALF*Pm*2	1.0478	1.6257	0.0330	
ALF*Ttr*	1.3407	2.3195	1.2315	1.1622

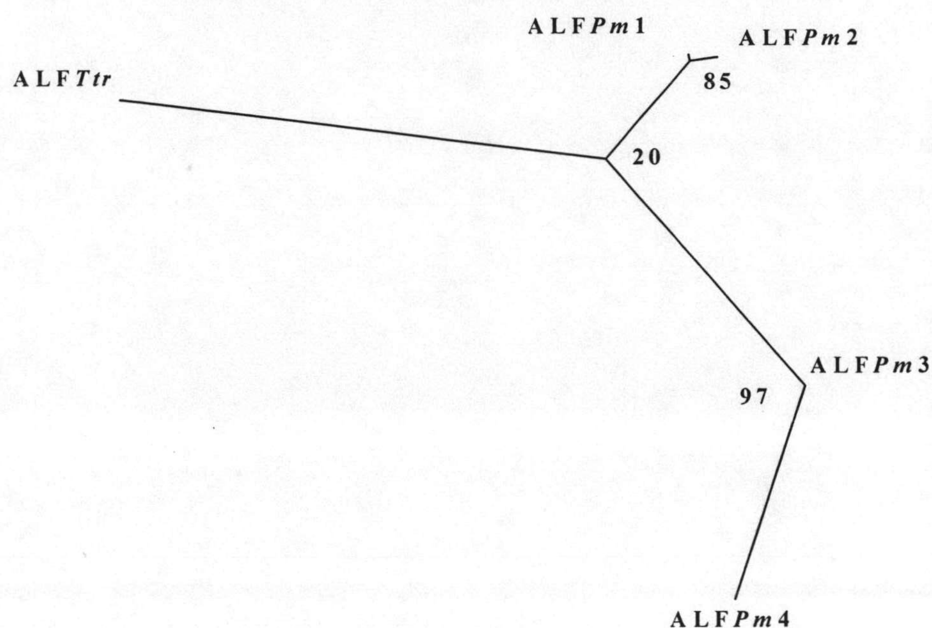

Figure 3.8 A bootstrapped NJ tree of phylogenetic tree of ALF genes found in the study. Values at the node represent the percentage of times that the particular node occurred in 500 tree generated by bootstraping of the original amino acid sequences.

Penaeidins

Penaeidins are members of a new family of antimicrobial peptides exhibiting Gram-positive antibacterial and antifungal activities and initially isolated from penaeid shrimp. Three putative peptides namely penaeidins2a, -2b and –2c were previously reported in *P. vannamei*. Thirteen penaeidin homologues were identified in *P. monodon* by the present study. Ten of these contain complete ORFs. Nucleotide sequences of representative sequences were aligned and showed high similarlity between these ESTs (Figure 3.9).

All ESTs contained an ORF of 222 bp coding for a protein of 74 amino acids. The putative signal peptide was predicted at NH_2-terminal composing of 19 amino acid residues that are almost identical to that of *P.vannamei* (Figure 3.10). The sequences after the putative signal peptide were composed of a proline-rich region followed by a COOH-terminal region containing 6 cysteine residues at conserved locations. Deduced amino acid sequences were aligned and suggested 3 different peptides which were different from one other by a single amino acid. These newly isolated sequences revealed the highest homology to Pen-2 of *P. vannamei* (Figures 3.9-3.10). They were then named as Pen*Pm*2a, Pen*Pm*2b, Pen*Pm*2C, respectively. The genetic distance (Table 3.11) and phylogenetic analysis indicated that putative penaeidins in *P. monodon* were more closely related within species (0.0136 – 0.0276) than between species (0.5713 – 0.6004) (Figure 3.11).

```
PenPm2b     ATGCGTCTCGTGGTCTGCCTGGTCTTCCTGGCCTCCTTCGCCCTGGTCTGCCAAGCCCAA
PenPm2c     ATGCGTCTCGTGGTCTGCCTGGTCTTCCTGGCCTCCTTCGCCCTGGTCTGCCAAGCCCAA
PenPm2a     ATGCGTCTCGTGGTCTGCCTGGTCTTCCTGGCCTCCTTCGCCCTGGTCTGCCAAGCCCAA
Pen-2       ATGCGCCTCGTGGTCTGCCTGGTCTTCTTGGCCTCCTTCGCCCTGGTCTGCCAAGGCGAA
            ***** ********************* *************************** * **

PenPm2b     GGGTACCAGGGTGGTTACACACGCCCGTTCCCCAGACCACCCTATGGGGAGGAT-ATCA
PenPm2c     GGGTACCAGGGTGGTTACACACGCCCGTTCCCCAGACCACCCTATGGGGAGGAT-ATCA
PenPm2a     GGGTACCAGGGTGGTTACACACGCCCGTTCCCCAGACCAACCTATGGGGAGGAT-ATCA
Pen-2       GCGTACAGGGGCGGTTACACAGGCCCGATACCCAGGCCACCACCCATTGGAAGACCACCG
            * ****  *** ********* ***** * ***** *** *       *** **   * *

PenPm2b     TCCAGTTCCTGTTTGTACTTCATGCCACAGGCTTAGCCCCTTACAAGCTCGTGCTTGCTG
PenPm2c     TCCAGTTCCTGTTTGCACTTCATGCCACAGGCTTAGCCCCTTACAAGCTCGTGCTTGCTG
PenPm2a     TCCAGTTCCTGTTTGCACTTCATGCCACAGGCTTAGCCCCTTACAAGCTCGTGCTTGCTG
Pen-2       TTCAGA-CCTGTTTGCAATGCATGCTACAGACTTTCCGTCTCAGATGCTCGCAATTGCTG
            * ***  ******** * * ***** **** ***   *  ** * * *****    ******

PenPm2b     CAGGCAGTTAGGACGTTGTTGTGATGCAAAGCAAACATATGGTTGA
PenPm2c     CAGGCAGTTAAGACGTTGTTGTGATGCAAAGCAAACATATGGTTGA
PenPm2a     CAGGCAGTTAG-----------------------------------
Pen-2       CATCAAGTTCGGAAGCTGTTGTCACTTAGTAAAAGGATAA------
            **   ****
```

Figure 3.9 Nucleotide sequence alignment of Pen*Pm*2a-Pen*Pm*2c of *P. monodon* with pen-2 of *P. vannamei*

```
PenPm2a     MRLVVCLVFLASFALVCQAQGYQGGYTRPFPRPTYGGGYHPVPVCTSCHRLSPLQARACC
PenPm2c     MRLVVCLVFLASFALVCQAQGYQGGYTRPFPRPPYGGGYHPVPVCTSCHRLSPLQARACC
PenPm2b     MRLVVCLVFLASFALVCQAQGYQGGYTRPFPRPPYGGGYHPVPVCTSCHRLSPLQARACC
Pen2        MRLVVCLVFLASFALVCQGEAYRGGYTGPIPRPPPIGRPPFRPVCNACYRLSVSDARNCC
            ******************.:.*:**** *:***.   *     ***.:*:***   :** **

PenPm2a     RQLGRCCDAKQTYG
PenPm2c     RQLRRCCDAKQTYG
PenPm2b     RQLGRCCDAKQTYG
Pen2        IKFGSCCHLVKG--
             ::  **.  :
```

Figure 3.10 Amino acid sequence alignment of Pen*Pm*2a-Pen*Pm*2c of *P. monodon* with pen-2 of *P. vannamei.* Putative signal peptide is underlined.

Table 3.11 Genetic distance calculated from amino sequence divergence of penaeidins in *P. monodon* and an outgroup (Pen2)

	Genetic distance		
Pen*Pm*2a	Pen*Pm*2a	Pen*Pm*2c	*PenPm2b*
Pen*Pm*2c	0.0275		
Pen*Pm*2b	0.0136	0.0136	
Pen2	0.6003	0.6003	0.5713

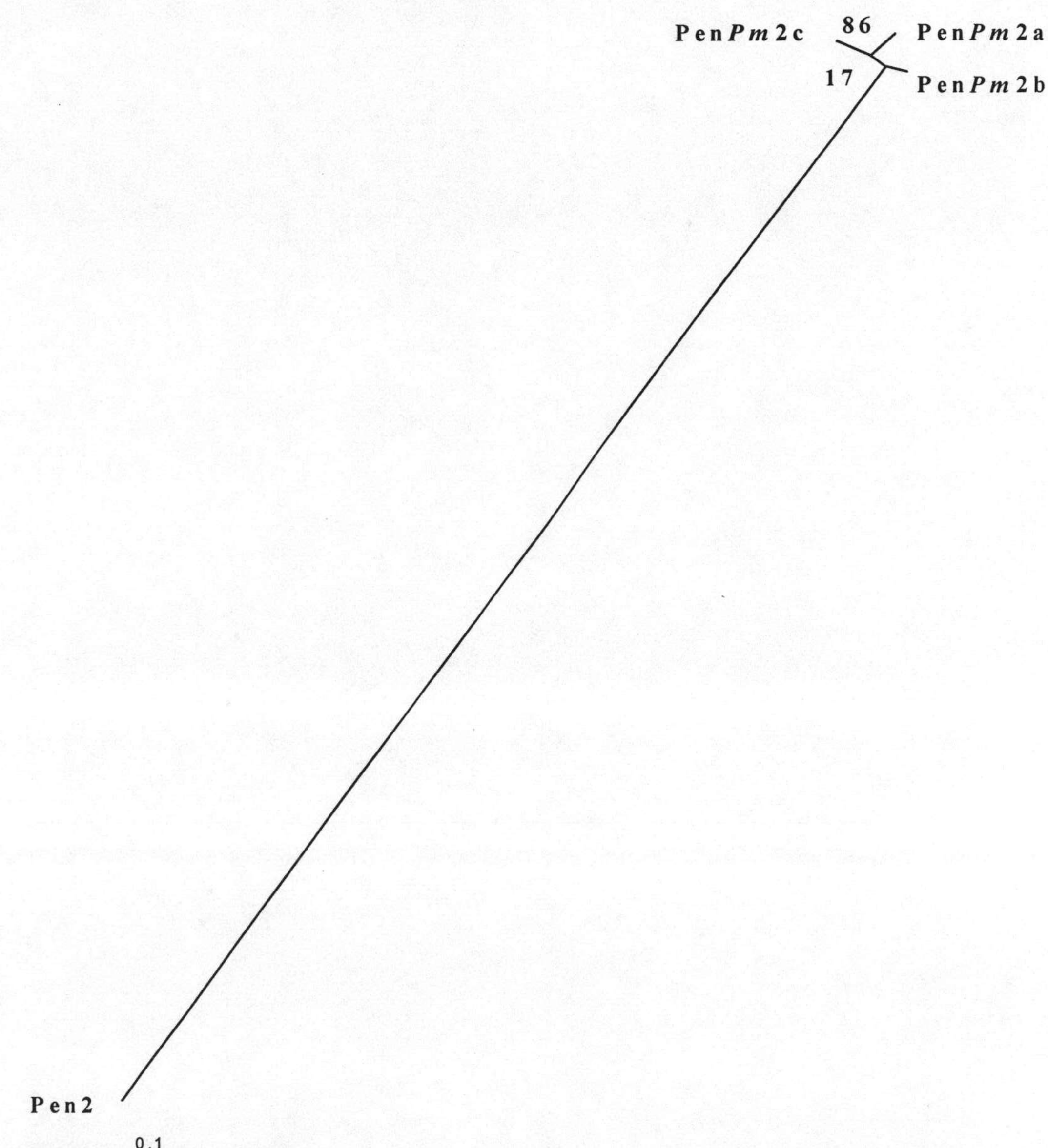

Figure 3.11 A bootstrapped NJ tree of phylogenetic tree of penaeidin genes found in the study. Values at the node represent the percentage of times that the particular node occurred in 500 tree generated by bootstraping of the original amino acid sequences.

Serine protinase inhibitors

Nine clones which were homologues of serine proteinase inhibitors of the Kazal family were identified. Four full-length clones obtained contained the ORFs of 804, 666 and 651 bp encoding for 268, 222 and 217 amino acid acids, respectively (Figure 3.12). From their deduced amino acids, 3 different proteins (SPI*Pm*1, SPI*Pm*2 and SPI*Pm*3) were recognized. These putative protinase inhibitors showed 50-65% homology to a four-domain Kazal inhibitor from crayfish, *Pacifastacus leniusculus* (Figure 3.13). A putative signal peptide at the NH_2-terminal was predicted (Figure 3.13). From the deduced amino acid sequence of *P. monodon* inhibitors, they differ in the number of the Kazal domain. Each domain contains 6 cysteine residues. SPI*Pm*1 and SPI*Pm*2 contain complete 5 and 4 Kazal domains, respectively whereas SPI*Pm*3 has 2 complete and 2 incomplete Kazal domains.

The genetic distance (Table 3.12) and phylogenetic tree (Figure 3.14) showed closed relationships between SPI*Pm*1-3 and crayfish where SPI*Pm*1 and SPI*Pm*3 showed the closest relationship and clustered together. SPI*Pm*3 and that of the crayfish were located in different branches of the tree.

Crustin

Thirteen clones are homologues of crustins, which inhibited the growth of Gram-postive bacteria, previously reported in *P. vanamei* and *P. sertiferus*. Two crustin homologue clones contained an identical complete ORFs (Crustin*Pm*1) of 435 bp encoding for a protein of 84 amino acids and con tained 19 residues of a signal peptide. Nucleotide sequences of Crustin*Ps* and Crustin*Pm*1 are highly similar (Figure 3.15). The deduced amino acid alignment suggested 5 different peptides (called crustin*Pm*1-crustin*Pm*5) (Figure 3.15). Crustin*Pm*3 was similar to crustin*Pm*1 and crustin*Pm*2 at domain I, whereas crustin*Pm*3 was more similar with crustin*Pm*4 and crustin*Pm*5 at domain II. The deduced amino acid of this protein showed 65%

```
SPIPm2     ------------------------------------------------------------
SPIPm3     -ATGGCCAACAAAGTGGCACTCTTGACCCTTCTTGCAGTGGCCGTTGCAGTCTCTGGCTA
SPIPm1     ATGCCTGAATTACCTGAGGTTGAGGTAACCCGTTTAGGAATTACCCCACATGTTTTACAT

SPIPm2     ---ATGTTGTTGTGCAAGATTACTCTTATCCATCTCCTGTTGCAAGGATTTGC-------
SPIPm3     CGGAAAGGGGGGAAAATCCGACTCTGCGCCAAACACTGTA-CGACCATCTCCCCTGTGT
SPIPm1     CAAACAGTAACGGATATTGTTATTCGTAATGGACGCTTACGTTGGCCAATTCCAGATGAT
              *       *   *     * **            *         *  * *

SPIPm2     -TGTCTTTAATGACGCCAACTCCGATCATGATTGTAT-------CGGCTACTGTCCTGAA
SPIPm3     GTGGCTCTGATGGAAAAACGTATGACAGCCGATGCCC-------AGGATTATGCCCCGCG
SPIPm1     ATTAATCAAATAAAGCAACAACCGATTACTAAAGTACGCCGTCGAGCGAAATATTTGTTG
            *   *   **      *     **        *           *     *

SPIPm2     GTGTATGATCCTGTGTGTGCCAGTAACGGC----TGGACTTACAACAACGACTGCGAACT
SPIPm3     GTATATGCCCCTGTGTGCGGGACCAACGGG----AAAACTTACTCGAACTTATGCCAACT
SPIPm1     TTAGATACGCCTGTTGGTAGTGCCATTGTGCATTTGGGGATGTCGGGAAGTTTGCGTGTT
            *  **   *****  *       *  *            *      *    ***    *

SPIPm2     ACAGGCTATGATAAAGTGCC---AGGGATGGAATATCACCAAGACACACGACCAAGCATG
SPIPm3     TGAGAATGACAGAACCTGCA---ACGGTGCTTTCGTTTCCAAGAAGCACGATGGACGTTG
SPIPm1     TTACCTGCTGGAACGGCTCCTGAAAAGCACGATCATGTTGATTTGGCATTATCGAGCGGG
             *         *     *    *  *        *    *      **  *   *    *

SPIPm2     TGAATGC---CTCAAGG---CCTGCCCCACGACCTTTGCCCCTGTGTGTGGG-------T
SPIPm3     TGGTTGCAACCCCAATGTTGCGTGCCCTGAGATCTATGCTCCCGTGTGTGGC-------A
SPIPm1     GAAATCCTACGTTATAA---CGATCCACGTCGTTTTGGGGCTTGGTTATGGCAACCTGTT
               * *      *      *   **        *  *  *  *  * ***

SPIPm2     CAGACAACAAGACCTATCTCAACGAGTGTGTCTTCGAGGTGGCTTCTTGCTGGGATCA-T
SPIPm3     GTGATGGCAAGACTTATGATAACGACTGCTATTTCCAGGCAGCTGTTTGCAAGAATCC-A
SPIPm1     GATACAAAACATCATGTCTTAGCGAAATTAGGTCCTGAGCCGTTAACCGATGTATTCACT
              *    *   * *   *  ***       **    *  * *    *       **

SPIPm2     TCGCTCGACAAGGCGTCTGAAGGAGCTTGTGGCTGGGGTATCCATTGCCTGCAGTACTGC
SPIPm3     GATCTTAAGAAGGTTCGAGACGGTAACTGCGACTG------CACTCCTCTCATCGGCTGT
SPIPm1     GCTGATTATTTACAGCAAAAAGCGAAGGGTAAGCG------CACCGCGATAAAGCAATTT
                  *           * *      *     *      *       *       *

SPIPm2     CCTGAGGTATACGACCCTGTGTGTGGAAGCAACGGCCAGACTTACACGAACGAATGCGAG
SPIPm3     CCCAAGAACTACAGGCCTGTGTGTGGCAGCGACGGTGTAACTTACAACAACGACTGCTTC
SPIPm1     ATTATGGACAATCACATTGTG-GTGG--GTGTTGGTAATATCTATGCCAATGAGTCACTG
                *    *      **** ****  *   **     *  **     ** ** *

SPIPm2     TTGCAGGCTGCCATACAGTGTCGTGGGTTGCAGATTG-CAAAGAGGCACGACCAGGCTTG
SPIPm3     TTCAAGGTTGCT---CAGTGCAAGAACCCCGCGCTCGTCAAAGTCTCTGATACTCGCT-G
SPIPm1     TTTTCGGCGGGTATCCATCCTCAAAAA--GCGGCTGG-AGAGGTGACTCCTCAAGCATTG
           **   **  *     **               * * *   * *   *          * *

SPIPm2     TGAGTGCCACGCCACTTGCCCCCTGATCCACGACCCTGTTTGTGGCACTGACGATAGGAC
SPIPm3     TGAATGCAACACGTCTGTACTGAAGAAT-ATTACCC-GTGTGCGGAAGC-ATGGTGTCAC
SPIPm1     ACTGTGCTGGTGGATGAAATTAAAGCTGTATTAGCCTTTGCTATTCAAC-AAGGTGGTAC
               ***                 *    *  * **  *       *   * * *   **

SPIPm2     TTACTACAACGAGTGCT----TCTTCAC--TAAAGCTTCTTGTTGGGAT----AGGTCCA
SPIPm3     GTATTCGAACATTTGTC----TGTTGAA--TAATGCAGCGTGTTTAGAT----TCCTCTA
SPIPm1     AACGCTAAAAGATTTTAAAAATGCTGATGGTAAACCTGGCTATTTTGCTCAAGAACTACA
                  **     *       *  * *   *** *    * **   * *         *  *

SPIPm2     TTTTGAAGAAGAAAAACGGACCTTGTGACAGGAAATGG--AAATACCTTCTGGAGATTTA
SPIPm3     TTTACAAGGTTTCGGACGGAATCTGTGGTCGCAA---G--ATGTACCTATAA--------
SPIPm1     AGTGTATGGTAAAGGTGGCCTGCCGTGCCCTAAGTGCGGTACTGAGCTTGCCGAGGTGAA
             *  * *         *      ***     *    *  *    * **

SPIPm2     A-----------------------------------------------
SPIPm3     ------------------------------------------------
SPIPm1     AATAGGCCAGAGGGCAACGGTTTATTGCTCACAGTGTCAGCAATAA
```

Figure 3.12 Multiple nucleotide sequence alignment of SPI*Pm*1-3 of *P. monodon*

```
SPIPle
AAR
CPSI—CPLNYKPVCGSDLKTYGNSCQLNAAICRNPSLKKLYDGP-----CIDKP
CPSI—CPLDYNPVCGTDGKTYSNLCALRIEACNNPHLNLRVDYQGE---CRP
CRNG—CTLQYDPKCGTDGKTYSNLCDLEVAACNNPQLNLKVAYKGE---CKQ
CPTI—CTQQYDPVCGTDGKTYGNSCELGVAACNNPQLNNKIAYKGA---CNF
PQQQT


SPIPm1
MANKVALLTLLAVAVAVSGYGKGGKIRL
CAKH--CTTIS-PVCGSDGKTYDSRCHLENAA-CGGVSVTFHHAGPCPPPKR
CPGI--CPAVYAPVCGTNGKTYSNLCQLENDRTCNGAFVSKKHDGRCG
CNPIVACPEIYAPVCGSDGKTYDNDCYFQAAV-CKNPDLKKVRDGNCD
CTPLIGCPKNYRPVCGSDGVTYNNDCFFKVAQ-CKNPALVKVSDTRCE
CNHV--CTEEYYPVCGSNGVTYSNICLLNNAA-CLDSSIYKVSDGICG
RRLYLZ

SPIPm2
MLLCKITLIHLLLQGFAVFNDANSDHD
CIGY--CPEVYDPVCASNGWTYNNDCELQAMIKCQGWNITKTHDQACE
CLKA--CPTTFAPVCGSDNKTYLNECVFEVAS-CWDHSLDKASEGACGWGIH
CLQY--CPEVYDPVCGSNGQTYTNECELQAAIQCRGLQIAKRHDQACE
CHAT--CPLIHDPVCGTDDRTYYNECFFTKAS-CWDRSILKKKNGPCD
RKWKYLLEI

SPIPm3
MANKVALLTLLAVAVAVSGYGKGGKIRL
CAKH--CTTIS-PVCGSDGKTYDSRCPGL----CPAVYA
            PVCGTNGKTYSNLCQLENDRTCNGAFVSKKH-DGRCG
CNPNVACPEIYAPVCGSDGKTYDNDCYFQAAV-CKNPDL-KKVRDGNCD
CTPLIGCPKNYRPVCGSDGVTYNNDCFFKVAQ-CKNPALV-KVSDTRCE
CLLNNAACLDSSIYKVSDGICGRKMYL
```

Figure 3.13 Deduced amino acids of proteinase inhibitors of *P. monodon* (SPI*Pm*1-3) and crayfish, *Pacifastacus leniusculus* (SPI*Ple*) were aligned. Kazal motifs were arranged. Putative signal peptide is underlined.

Table 3.12 Genetic distance calculated from amino sequence divergence of SPI in *P. monodon* and an outgroup (SPI*Ple*)

	Genetic distance		
SPI*Pm*1	SPI*Pm*1	SPI*Pm*3	SPI*Pm*2
SPI*Pm*3	0.0762		
SPI*Pm*2	1.0101	1.0553	
Crayfish	1.0950	1.2116	1.4087

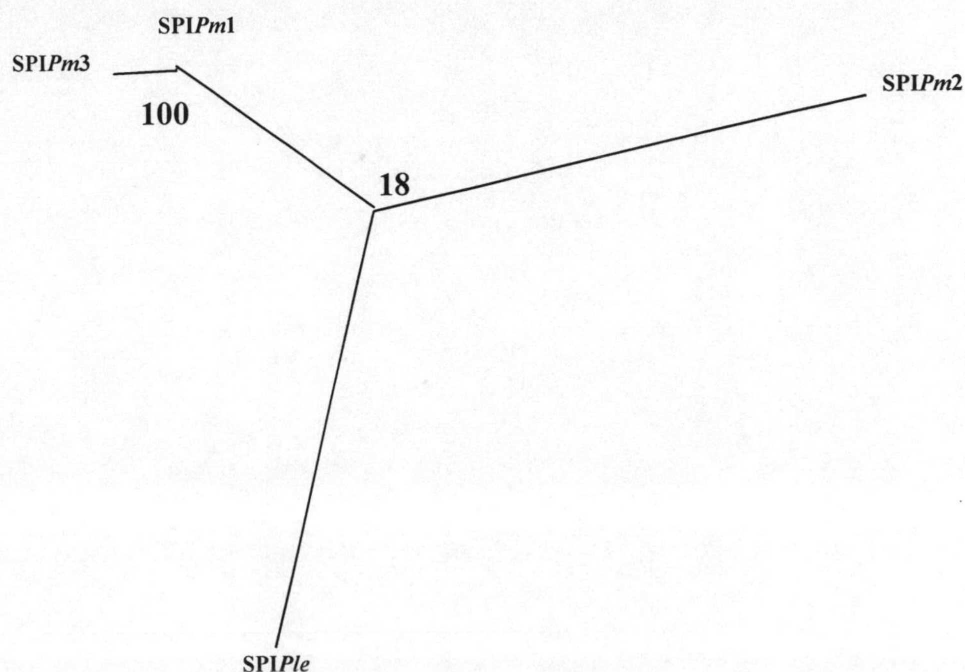

Figure 3.14 A bootstrapped NJ tree of phylogenetic tree of SPIs genes found in the study. Values at the node represent the percentage of times that the particular node occurred in 500 tree generated by bootstraping of the original amino acid sequences.

homology with that of *P. setiferus* (Figure 3.16). The putative signal peptide was predicted at the NH_2-terminal and located between ala (A) and gln (Q) (Figure 3.15a). The protein contains glycine-rich repeat region at NH_2-terminal and 2 domains of cysteine-rich residues. A 4-disulfide core (4-DSC) domain from 8 cystein residues was found in conserved arrangement at Domain II. This domain called Whey acidic protein (WAP). Domain II of the WAP appeared more conserved than did Domain I.

The genetic distance (Table 3.13) and phylogenetic tree (Figure 3.17) illustrated that Crustin*Pm*1 and Crustin*Pm*2 were closely related to that of *P. sertiferus*. The closest relationship was found between Crustin*Pm*4 and Crustin*Pm*5 whereas Crustin*Pm*3 was allocated into a separate branch.

Heat shock protein 10

A clone from the normal library coding a homologue of a small heat shock protein (HSP10) contained the complete ORF of 309 bp, encoding a 102 amino acid protein. The nucleotide sequence of this gene homologue was aligned with that previously reported in vertebrates (Figure 3.18). Deduced amino acid of *P. monodon* heat shock protein 10 showed 73 to 75% homology with those of vertebrates, *Mus Musculus*, *Homo sapiens* and *Gallus gallus* (Figure 3.19). These proteins showed the amino-terminal conserved motif of cnp10, PXX(D/N)(K/R) (bold face).

Large genetic distance (Table 3.14) was observed between heat shock protein 10 of *P. monodon* and that of others (0.6870 – 0.0.7598) while much lower genetic distance was observed between heat shock protein of *H. sapiens*, *M. musculus* and *G. gallus* (Figure 3.20). A phylogenetiv tree revealed clear differentiation between *P. monodon* heat shock protein 10 and that of 3 vertebrate species.

```
CrustinLs     ATGAAGGGTCTCGGAGTCATTCTGTGCTGCGTCCTGGCGGTGGTTCCAGCCCAC-----G
CrustinPm1    ATGAAGGGTCTCGGAGTCATTCTGTTCTGCGTCCTGGCGATGGCATCAGCCCAGAGTTGG
              ************************* ************* ***   *******      *

CrustinLs     CAGGG-------CCCGGAGGCTTCTCTGGTGGCGTT---CCCGGAGGCTTCCCTGGTGGA
CrustinPm1    CACGGAGGTCGACCCGGAGGCTTCCCTGGTGGAGGTAGACCCGGAGGCTTCCCTGGTGGA
              ** **       ************ ******* * *   *********************

CrustinLs     CGTCCCGGAGGCTTCCCTGGTGGCGTTCCCGGAGGCTTCCCCAGCGCCACAGCTCCTCCC
CrustinPm1    GGT-----AG----ACCCGGAGGTAGACCCGGAGGCTTCCCAAGCGTCACAGCCCCACCC
               **     **     ** ** **    ************** **** ****** ** ***

CrustinLs     GCCACATGCAGGCGCTGGTGCAAAACTCCGGAGAATCAAGCCTACTGCTGCGAGACCATC
CrustinPm1    GCCTCCTGTAGGCGCTGGTGCGAAACTCCAGAGAATGCTTTTTACTGCTGCGAGTCAAGG
              *** * ** ************ ******* ******      ************* * *

CrustinLs     TTTGAACCCGAGGCCCCCGTGGGCACCAAGCCCCTTGACTGCCCACAAGTCCGTCCCACC
CrustinPm1    TATGAACCCGAGGCACCCGTGGGCACCAAGATACTTGACTGCCCAAAAGTCCGTGACACC
              * ************ ***************   ************ ********  ****

CrustinLs     TGCCCACCCACTCGCTT-----CGGTGGA-CGGCCTGTAACCTGCTCCAGCGACTACAAG
CrustinPm1    TGCCCACCGGTACGTTTTCTTGCAGTAGAGCAGCCAGTACCTTGCTCCAGTGACTACAAG
              ********    ** **     * ** ** * *** *** * ******** *********

CrustinLs     TGTGGCGGCCTGGACAAGTGCTGCTTCGACAGGTGTCTGGGAGAACACGTGTGCAAGCCC
CrustinPm1    TGCGGCGGCCTTGACAAGTGCTGCTTCGACAGGTGTTTGGGACAACACGTGTGCAAGCCA
              ** ******** ************************ ***** ****************

CrustinLs     CCTTCCTTCTACAGCCAATTCCGTTGA
CrustinPm1    CCTTCCTTCTATGAATTTTTGCCTGA
              ***********       **    ***
```

(a)

```
crustinLs     MKGLGVILCCVLAVVPAHAGPGGFSGGVPGG-FPGGRPGG-FPGGVPGGFPSATAPPATC
crustinPm1    MKGLGVILFCVLAMASAQSWHGGRPGGFPGGGRPGGFPGGGRPGGRPGGFPSVTAPPASC
              ******** ****:..*::  ** .**.***  *** ***  *** ******.*****:*

crustinLs     RRWCKTPENQAYCCETIFEPEAPVGTKPLDCPQVRPTCPPTRFGG--RPVTCSSDYKCGG
crustinPm1    RRWCETPENAFYCCESRYEPEAPVGTKILDCPKVRDTCPPVRFLAVEQPVPCSSDYKCGG
              ****:****  ****: :********* ****:** ****.** .  :**.*********

crustinLs     LDKCCFDRCLGEHVCKPPSFYSQFR
crustinPm1    LDKCCFDRCLGQHVCKPPSFYEFFA
              ***********:*********. *
```

(b)

Figure 3.15 Multiple sequence alignment of nucleotide (a) and deduced amino acid (b) of crustin*Pm*1of *P. monodon* with a crustin of *P. setiferus*. Putative signal peptide is underlined.

```
CrustinPs         ------------------------------------------------------------
CrustinPm1        ------------------------------------------------------------
CrustinPm2        ------------------------------------------------------RFRHEA
CrustinPm4        ------------------------------------------------------------
CrustinPm5        ---------------------------------------------RRRIRRRFRRRCPRR
CrustinPm3        GGGAYGGGLGGGLGGGGVNGGGLGGGLGGGVNGGGLGGGLGGGVNGGGLGGGLGGGVHGG

CrustinPs         MKGLGVILCCVLAVVPAHAGPGGFSGGVPGGFPGG--------RPGGFPGGVPGGFPSAT
CrustinPm1        MKGLGVILFCVLAMASAQSWHGGRPGGFPGGGRPGGF------PGGGRPGGRPGGFPSVT
CrustinPm2        SXFTGFRSHSFXKHLAVVSAHGGRPGARPGGFPAGVPGGFPGGVPGEFPAPHLGGFLSVT
CrustinPm4        AVXEAVSRRRFRRRFRRRCPRRRLRRRFRRRFKWWFG----WWSPSSGRWSRSSGWWSKT
CrustinPm5        RLRRRFRRRCQRRRFRRRCPRRRLRRRFRRRFKWWFG----WWSRSSGWWSRSSGWWSKT
CrustinPm3        GLGGGLGGGVNGGGLGGGVHGGGLGGGLGGGLSGGLGGG--LGRPGGGLGRPGGGLRPGS
                       .                                        .       .* . :

                  __________DOMAIN I____________   ____________DOMAIN II_____
CrustinPs         APPATCRRWCKTPENQ---AYCCETIFEPEAPVGTKPLDCPQVRPTCPPTRFG--GRPVT
CrustinPm1        APPASCRRWCETPENA---FYCCESRYEPEAPVGTKILDCPKVRDTCPPVRFLAVEQPVP
CrustinPm2        APPATCRRWCRTPEDA---VYCCESKYEPEAPVGTKPLDCPRVRDTCPPVRFGG-LAPVT
CrustinPm4        GFSWAKHFQVLVHHSG-GKQYCCEDKNEPEIPVGTKPLDCPQIRPTCPRFQRP----PVT
CrustinPm5        GFSWAKHLQVLVHHSGRVNSTAARIRMEPEIPVGTKPLDCPQVRPTCPRFQGP----PVT
CrustinPm3        RGPSTCRYWCTTPEGK---QYCCEDKNEPEIPVGTKPLDCPQVRPTCPRFQGP----PVT
                     . : :  . ..       ...   *** ***** ****::* ***  :      **.

                  ________________________
CrustinPs         CSSDYKCGGLDKCCFDRCLGEHVCKPPSFYSQFR--
CrustinPm1        CSSDYKCGGLDKCCFDRCLGQHVCKPPSFYEFFA--
CrustinPm2        CSSDLKCGGLDKCCFDRCLKEHVCKPPSFYSHFA--
CrustinPm4        CSHDFKCAGLDKCCFDRCFGEHVCQPPSFYGRNVKG
CrustinPm5        CSHDFKCAGLDKCCFDRCLGEHVCKPPSFYGRNVKG
CrustinPm3        CSHDFKCAGLDKCCFDRCLGEHVCKPPSFYGRNVKG
                  ** * **.**********: :***:*****
```

Figure 3.16 Sequence alignment of deduced amino acid of crustin*Pm*1-5 of *P. monodon*

Table 3.13 Genetic distance calculated from amino sequence divergence of crustins in *P. monodon*. Crustin previously found in the *P. sertiferus* (crustin*Ls*) use as an outgroup.

	Genetic distance				
Crustin*Ls*	Crustin*Ls*	Crustin*PM*1	Crustin*PM*2	Crustin*PM*4	Crustin*PM*5
Crustin*PM*1	0.3341				
Crustin*PM*2	0.4478	0.4602			
Crustin*PM*4	1.0667	1.2296	1.1421		
Crustin*PM*5	1.1312	1.2409	1.2472	0.1949	
Crustin*PM*3	0.6171	0.7891	0.7548	0.7622	0.9829

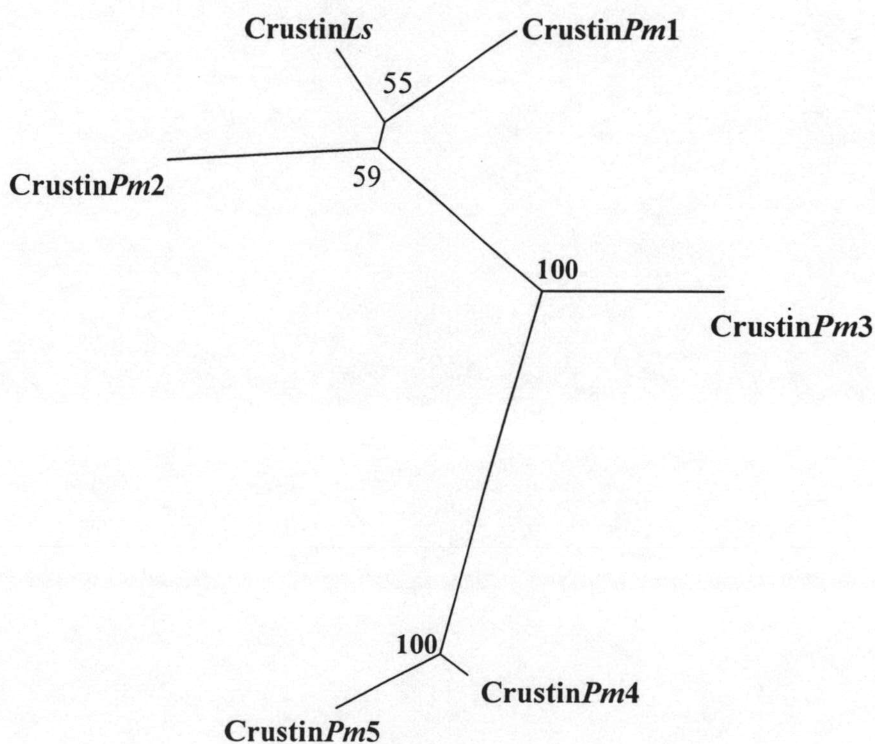

Figure 3.17 A bootstrapped NJ tree of phylogenetic tree of crustin genes found in the study. Values at the node represent the percentage of times that the particular node occurred in 500 tree generated by bootstraping of the original amino acid sequences.

```
Hsp10Hs         -------ATGGCAGGACAAGCGTTTAGGAAGTTTCTTCTGATTGCTGACAGGGTATTGGT
Hsp10Mm         -------ATGGCAGGACAAGCGTTTAGGAAGTTTCTTCTGATTGCTGACAGGGTATTGGT
Hsp10Gg         -------ATGGCAGGAAAAGCATTTAGGAAATTCCTTCCCCTGTTTGATCGTGTTCTGGT
Hsp10Pm         AGTAAAAATGGCTGGT---GCTTTGAAGAAGTTTGTTCCCTTGTTCGACCGTGTGCTGGT
                       ***** **    ** ** * *** **  ***   *    **  * **  ****

Hsp10Hs         TGAAAGGAATGCCGCCGAAATTGTAGCCAAAGGTGGCATTATGCTTCCAGAAAAGTCTCA
Hsp10Mm         TGAAAGGAATGCCGCCGAAATTGTAGCCAAAGGTGGCATTATGCTTCCAGAAAAGTCTCA
Hsp10Gg         TGAACGATGCGCAGCAGAGACCGTAACCAAAGGAGGCATCATGATTCCAGAAAAAGCTCA
Hsp10Pm         CCAAAAGGCTGAGGCTTTGACCCGTACTGCAAAGGGAATCTTGATCCCAGAAAAGTCTGT
                  **      *  **    *      *   *   ** **  ** * ********  **

Hsp10Hs         AGGAAAAGTATTGCAAGGAACGGTCGTGGCTGTGGGGTCAGGCAGGAAAGGAAAGAGTGG
Hsp10Mm         AGGAAAAGTATTGCAAGGAACGGTCGTGGCTGTGGGGTCAGGCAGGAAAGGAAAGAGTGG
Hsp10Gg         AGGGAAAGTGCTGCAAGCAACAGTGGTAGCAGTTGGATCGGGAGCCAGAGGGAAGGATGG
Hsp10Pm         ACCCAAGGTCCTCACAGGGAAGGTGGTTGCTGTAGGAGAGGGAGCCAGAACTGATGCTGG
                *   ** **  *   **  *  ** ** ** ** **    **    * *    *   ***

Hsp10Hs         AGAGATTGAGCCTGTCAGTGTGAAAGTTGGAGATAA------------------------
Hsp10Mm         AGAGATTGA---------------------------------------------------
Hsp10Gg         TGAGATTCATCCAGTGAGTGTAAAAGTTGGTGAAAAGGTTTTGCTACCAGAATATGGTGG
Hsp10Pm         CACCACAATCCCCCCATGTGTTACTGTTGGTGATGAAGTGATGCTTCCTGAGTTTGGTGG
                    *

Hsp10Hs         ------------------------------------------------------------
Hsp10Mm         ------------------------------------------------------------
Hsp10Gg         TACCAAAATTGTACTAGAAGATAAGGACTACTACTTGTTTAGAGACGGTGACATTCTTGG
Hsp10Pm         CACAAAGGTTACCCTGGAGGAGAAGGACTATTACCTCTTCAGAGAATCTGAACTCCTGGC

Hsp10Hs         -------------------
Hsp10Mm         -------------------
Hsp0Gg          AAAATACCTGGACTGA---
Hsp10Pm         CAAGATGAAGAACGAGTAA
```

Figure 3.18 Nucleotide sequence alignment of *P. monodon* heat shock protein 10 homologue with other mammalian HSP10; *Hs*= *Homo sapiens*, *Mm*=*Mus musculus*, *Gg*=*Galus galus* (Dickson et al., 1994).

```
Hsp10Mm         MAGQAFRKFLPLFDRVLVERSAAETYTKGGIMLPEKSQGKVLQATVVAVGSGGKGKSGEI
Hsp10Hs         MAGQAFRKFLPLFDRVLVERSAAETVTKGGIMLPEKSQGKVLQATVVAVGSGSKGKGGEI
Hsp10Gg         MAGKAFRKFLPLFDRVLVERCAAETVTKGGIMIPEKAQGKVLQATVVAVGSGARGKDGEI
Hsp10Pm         MAG-ALKKFVPLFDRVLVQKAEALTRTAKGILIPEKSVPKVLTGKVVAVGEGARTDAGTT
                *** *::**:********::. * * *  **::***:  *** ..*****.*.: . *

Hsp10Mm         EPVSVKVGDLKVLLPEYGGTKVVLDDKDYFLFRDSDILGKYVD-
Hsp10Hs         QPVSVKVGD-KVLLPEYGGTKVVLDDKDYFLLRDGDILGKYVD-
Hsp10Gg         HPVSVKVGE-KVLLPEYGGTKIVLEDKDYYLFRDGDILGKYLD-
Hsp10Pm         IPPCVTVGD-EVMLPEFGGTKVTLEEKDYYLFRESELLAKMKNE
                 * .*.**: :*:***:****:.*::***:*:*:.::*.*   :
```

Figure 3.19 Sequence alignment of deduced amino acids of *P. monodon* heat shock protein 10 homologue with other mammalian Hsp10; *Hs*= *Homo sapiens*, *Mm*=*Mus musculus*, *Gg*=*Galus galus* (Dickson et al., 1994). The amino-terminal conserved motif of Hsp10 was in bold face (PXX(D/N)(K/R)).

Table 3.14 Genetic distance calculated from amino acid sequence divergence of HSP10 of *P. monodon* and other vertetrates; *Hs*= *Homo sapiens*, *Mm*=*Mus musculus*, *Gg*=*Galus galus*.

	Genetic distance		
Hsp10*Hs*	Hsp10*Hs*	Hsp10*Mm*	Hsp10*Gg*
Hsp10*Mm*	0.0831		
Hsp10*Gg*	0.1520	0.1641	
Hsp10*Pm*	0.7597	0.7348	0.6870

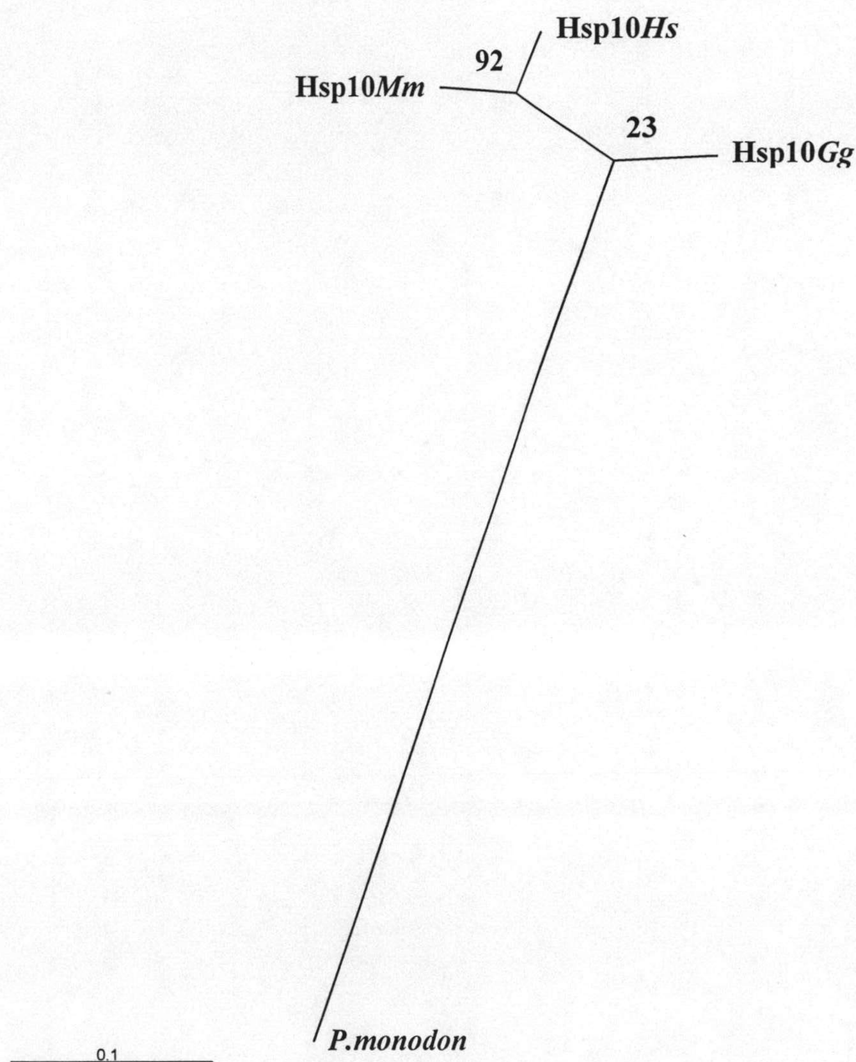

Figure 3.20 A bootstrapped NJ tree of phylogenetic tree of HSP10 genes from *P. monodon* and that of *M. muusculus*, *H. sapiens* and *G. gallus*. Values at the node represent the percentage of times that the particular node occurred in 500 tree generated by bootstraping of the original amino acid sequences.

Cytosolic manganese superoxide dismutase

Cytosolic manganese superoxide dimutase (MnSOD) is an enzyme that reduced the superoxide radical O_2^- into hydrogen peroxide and molecular oygen. In this study, a clone containing a complete ORF was identified from the infected library and subsequently clarified to be a homologue of cytosolic manganese superoxide dismutase. It is composed of 819 bp, coding a 273 amino acid protein which showed 86% homology with that of blue crab, *Callinectes sapidus* (Figure 3.21). Multiple amino acid sequence alignment of of *P. monodon* MnSOD and that of *C. sapidus* and *Ganoderma microsporum* (Figure 3.21) showed 4 conserved residues known to be involved in metal binding (His-112, His-159, Asp-244 and His-248) (Figure 3.22). The consensus sequence DXWEH located between Asp-244 and His-248. MnSODPm contained a puitative N-glycosylation site (NHT/S) within the molecule. The genetic distance (Table 3.15) and phylogenetic tree (Figure 3.23) revealed close relationship between MnSOD*Pm* and MnSOD*Cs* but distantly related relationships were observed when compare MnSOD*PM* and MnSOD*Cs* with MnSOD*Gm*.

```
MnSODCs         AATGGCAGAG---AAGGATCTATACATTGCTGCCCTTGAGAAGAAGCTGGCTGAGCTGTC
MnSODPm         -ATGGCTGAGGCAAAGGAAGCTTACATCTCCATCCTGGAGAAGAAGTTAGCTGAGCTGAC
                 ***** ***   *****    *****  *   *** ********* * ********* *

MnSODCs         TGGTATTGAAGTTGATCAGATCAAGAAAAACCAGCTAGCCAATGCTTCAAGTGAGGCACG
MnSODPm         TGGAATTGAGGTGGATCAGATCAAGAAGAATCAGTTCGCAAATGCAGCAGATGAGGCTGT
                *** ***** ** ************** ** *** * ** *****  **  ******

MnSODCs         TTCCATTCGTGAGATGGCTGAGTACGTGGAGGGCATCCAGGTAAAGCAAGCTGGACAGGT
MnSODPm         CGCCATCCGTGAGATGGCTTCATATGTAGAGGGGATTGTTGTACAGCAGGCTGGTGTTGC
                  **** ************   ** ** ***** **    *** **** *****    *

MnSODCs         TATTACTGGTCAGGTGAATCCTCAGGTGGCTGCCATGTTTTCCCATATTAAGGCGGAGCT
MnSODPm         TCAGGCTGGTACAGTCAGTCCTCAGATTGCACAGATGTTTGCCCATATCAATGCTGAATT
                *    *****   ** * ******* * **    ****** ******* ** ** **  *

MnSODCs         AGGTGAGGAGCGTGGAGTACATTCCCTGCCACCTCTTGGATATGACTATGGTGCCCTGGA
MnSODPm         GGGTGAGGAACGAGGTGCTCATGCTTTGCCGCCTCTCAAGTATGATTTCAATGCCCTTGA
                 ******** ** ** *  *** *  **** *****    ***** *    ****** **

MnSODCs         ACCCCATATCTGCACCACCATCATGCAGATCCATCACACCAAACACCATCAAGGATACAT
MnSODPm         ACTCCACATCTCTGGCATGATCATGGAGATCCACCACACAAAGCATCACCAGGGCTACAT
                ** *** ****    **  ****** ******* ***** ** ** ** ** ** *****

MnSODCs         CAACAACCTGAAGGCAGCTGTAGAAAAGCTTACAGAAGCAGAGAAGGCTAATGATATTGG
MnSODPm         TAACAACCTAATTGCTGCTACAAAGAAGTTGGTGGAGTCAGAGGCTGCCAATGATGTGAA
                 ******** *  ** ***  * * *** *    **  *****   ** ****** *

MnSODCs         TGCTATGAATGCACTCCTTCCGGCCATCAAGTTCAATGGAGGTGGCCATTTGAACCATAC
MnSODPm         TGCAATGAATGCCCTTCTACCAGCTATCAAGTTCAATGGAGGTGGCCACTTGAACCACAC
                *** ******** ** ** ** ** *********************** ******** **

MnSODCs         AATCTTCTGGACTAACATGGCTCCAGGAGCTGGAGGAGAACCCTCAGGATCCATTGCTGA
MnSODPm         CATCTTCTGGACCAACATGGCTCCTGATGCTGGTGGTGAGCCAGAAGGAGCAATTGCACA
                 *********** *********** *  ***** ** ** **   **** * *****  *

MnSODCs         GATGATCAACAGGGATTTTGGCTCATTCCAGTCCTTCAAGGAAAAGTTCTCAGGTGCCAG
MnSODPm         AGCCATTGATGATAGCTTTGGATCATTCCAGTCCTTTAAGGACAAATTTTCTGCTGCCAG
                    **  *       ***** ************** ***** ** ** ** * ******

MnSODCs         TGTTGCTGTGAAAGGATCTGGCTGGGGTTGGCTTGGCTATTGCCCTAAGGATGACAAGCT
MnSODPm         CGTTGGAGTGAAAGGCTCTGGCTGGGGATGGCTCGGGTATTGCCCCAATAACAACAAGCT
                 ****  ******** *********** ***** ** ******** **  *  *******

MnSODCs         TGCTGTTGCCACCTGCCAGAACCAGGATCCCCTGCAGATCACCCATGGTCTTGTGCCATT
MnSODPm         TGAGATCGCCACTTGCCAGAACCAGGATCCCTTGCAGATCACTCATGGCCTGGTTCCATT
                **   * ***** ****************** ********** ***** ** ** *****

MnSODCs         GCTGGGTTTGGATGTGTGGGAGCATGCCTATTATCTCCAGTATAAGAACCTGCGTGCCGA
MnSODPm         GCTCGGTCTTGATGTCTGGGAGCATGCTTACTACCTTCAGTACAAGAACCTCCGTGCAGA
                *** *** * ***** *********** ** ** ** ***** ******** ***** **

MnSODCs         TTATGTTAAAGCCTTCTTCAATGTGATCAACTGGGCCAACGTGAATGAGCGTTTTGAAGC
MnSODPm         TTACGTGAAGGCCTTCTTCAATGTCATCAACTGG-CCGATGTGA----------------
                *** ** ** ************** ********* ** * ****

MnSODCs         AGCTCGTAAGGCAGCTGGACACTGA
MnSODPm         -------------------------
```

Figure 3.21 Nucleotide sequence alignment of *P. monodon* cytosolic manganese superoxide dismutase (MnSOD*Pm*) homologue with those of *C. sapidus* (MnSOD*Cs*) and *G. micrsporum* (MnSOD*Gm*).

```
MnSODCs   1  MAE-KDLYIAALEKKLAELSGIEVDQIKKNQLANASSEARSIREMAEYVEGIQVKQAGQV
MnSODPm   1  MAEAKEAYISILEKKLAELTGIEVDQIKKNQFANAADEAVAIREMASYVEGIVVQQAGVA
MnSODGm      ------------------------------------------------------------

MnSODCs  60  ITGQVNPQVAAMFSHIKAELGEERGVHSLPPLGYDYGALEPHICTTIMQIHHTKHHQGYI
MnSODPm  61  QAGTVSPQIAQMFAHINAELGEERGAHALPPLKYDFNALELHISGMIMEIHHTKHHQGYI
MnSODGm   1  ------------------------MAHVLPDLPYAYNALEPFISQQIMELHHKKHHQTYV
                                      .* ** * * :.*** .*.  **::**.**** *:

MnSODCs 120  NNLKAAVEKLTEAEKANDIGAMNALLPAIKFNGGGHLNHTIFWTNMAPG-------AGGE
MnSODPm 121  NNLIAATKKLVESEAANDVNAMNALLPAIKFNGGGHLNHTIFWTNMAPD-------AGGE
MnSODGm  37  NSLNAAEQAYAKASTPKER---IALQSALKFNGGGHINHSLFWKNLAPAKSEGKGNGGAL
             *.* ** :  .::. .::     ** .*:*******:**::**.*:**        .*.

MnSODCs 173  PSGSIAEMINRDFGSFQSFKEKFSGASVAVKGSGWGWLGYCPKDDKLAVATCQNQDPLQI
MnSODPm 174  PEGAIAQAIDDSFGSFQSFKDKFSAASVGVKGSGWGWLGYCPNNNKLEIATCQNQDPLQI
MnSODGm  94  ADGPLKSAIEQNWGSVDNFIKEFNATTAAIQGSGWGWLGLNPATKRLWITTTANQDPL-L
             ..*.: . *: .:**.:.* .:*..:..:::********  *  .:* ::*  ***** :

MnSODCs 233  THGLVPLLGLDVWEHAYYLQYKNLRADYVKAFFNVINWANVNERFEAARKAAGH
MnSODPm 234  THGLVPLLGLDVWEHAYYLQYKNLRADYVKAFFNVINWPM--------------
MnSODGm 153  SH--VPIIGVDIWEHAFYLQYLNVKADYLAAIWIVINFKEAERRLIEATK----
             :*  **::*:*:****:**** *::***: *:: ***:
```

Figure 3.22 Sequence alignment of deduced amino acids of *P. monodon* cytosolic manganese superoxide dismutase (MnSOD*Pm*) homologue with those of *C. sapidus* (MnSOD*Cs*) and *G. micrsporum* (MnSOD*Gm*). The conserved (DXWEH) is underlined. The residues involved in metal binding were in bold face.

Table 3.15 Genetic distance calculated from amino acid sequence divergence of MnSOD of *C. sapidus* (*Cs*), *G. microsporum* (*Gm*) and *P. monodon*.

	Genetic distance	
MnSOD*Cs*	MnSOD*Cs*	MnSOD*Pm*
MnSOD*Pm*	0.2566	
MnSOD*Gm*	0.8931	0.8206

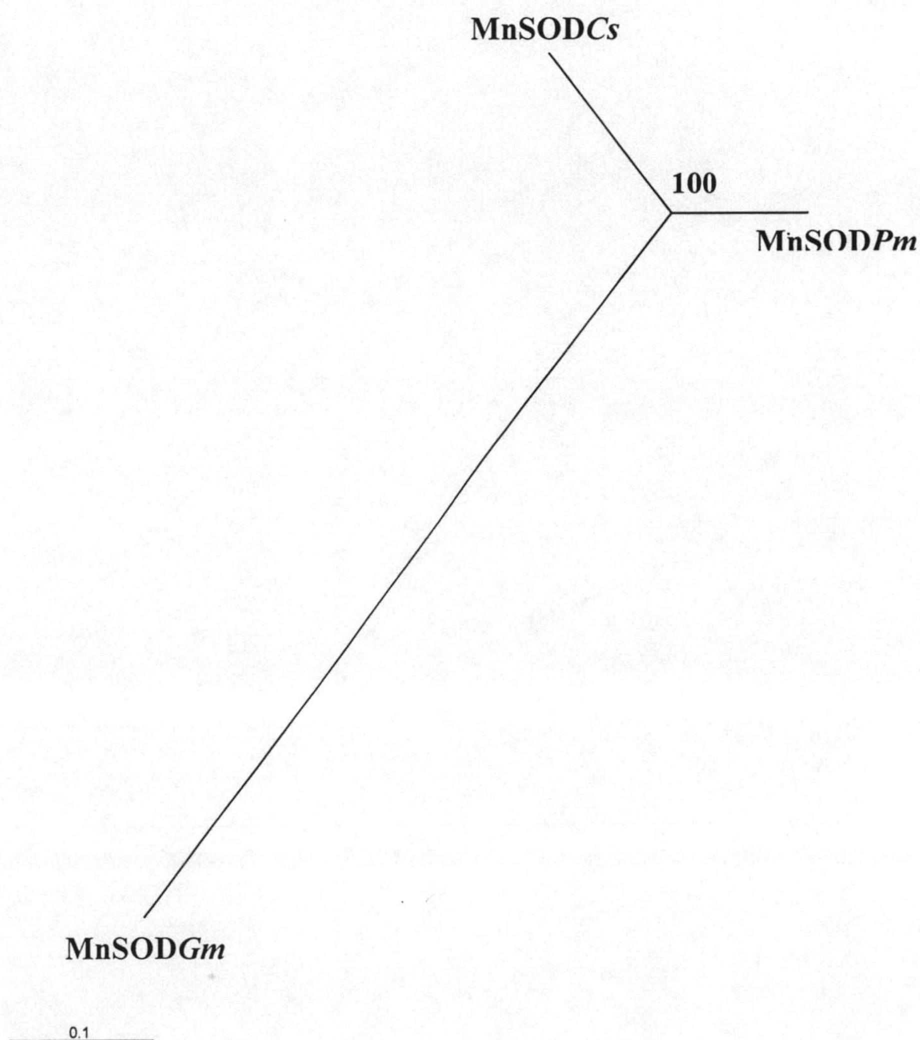

Figure 3.23 A bootstrapped NJ tree of the MnSOD gene found in this study and this of C. *sapidus*, *G. microsporum* and *P. monodon*. Values at the node represent the percentage of times that the particular node occurred in 500 tree generated by bootstrapping of the original amino acid sequences.

3.5 Simi-quantitative analysis of immune-related gene expres

Expression levels of 8 genes including ALF, crustin, penaeidin, HSP 90, HSP 70, lysozyme, serine protease inhibitor and prophenol oxidase, isolated by this study were examined by a semiquantitative RT-PCR. This technique requires the optimization of several parameters involving with PCR amplification to ensure a semiquantitative analysis of the transcripts.

3.5.1 Primer selection and determination of melting temperature (Tm)

Eight primer sets were designed for amplification of interesting genes based on the nucleotide sequence of the EST clones. In addition, primers for β–actin were designed as used as an internal control in a semi-quantitative PCR. The annealing temperatures were calculated for all primers and Tms were 50-60 °C. The annealing temperature at 53 °C was initially chosen. The amplification products of these genes were 250, 254, 216, 220, 419, 184, 170 and 217 bp for ALF, serine protinase inhibitor, crustin, penaeidin, prophenoloxidase, HSP 90, HSP 70 and lysozyme, respectively. Therefore, these transcripts were not size-overlapped with a 317 bp of β–actin amplification product. Non-specific products were not observed for all amplifications. Therefore, these primer sets were used to determine PCR condition for identifying the exponential phase of amplification at the annealing temperature of 53°C.

3.5.2 Determination of the optimal $MgCl_2$ concentration

Determination of the optimal concentration for each primer set was performed using different $MgCl_2$ concentrations (1-5 mM) using the standard PCR reaction. After the PCR products were run on an agarose gel, the concentration of $MgCl_2$ that gave the highest yields for each product was chosen (Table 3.16). ALF and serine protiase inhibitor amplification products gave the highest yield at 3 mM and 2 mM, respectively (Figure 3.24 a and b), whereas the remaining genes gave the highest yields at 1.5 mM $MgCl_2$ (Figure 3.24 c-h). An internal control, β–actin, showed the highest amplification product at all $MgCl_2$ concentrations (1 to 5 mM) (Figure 3.24 i).

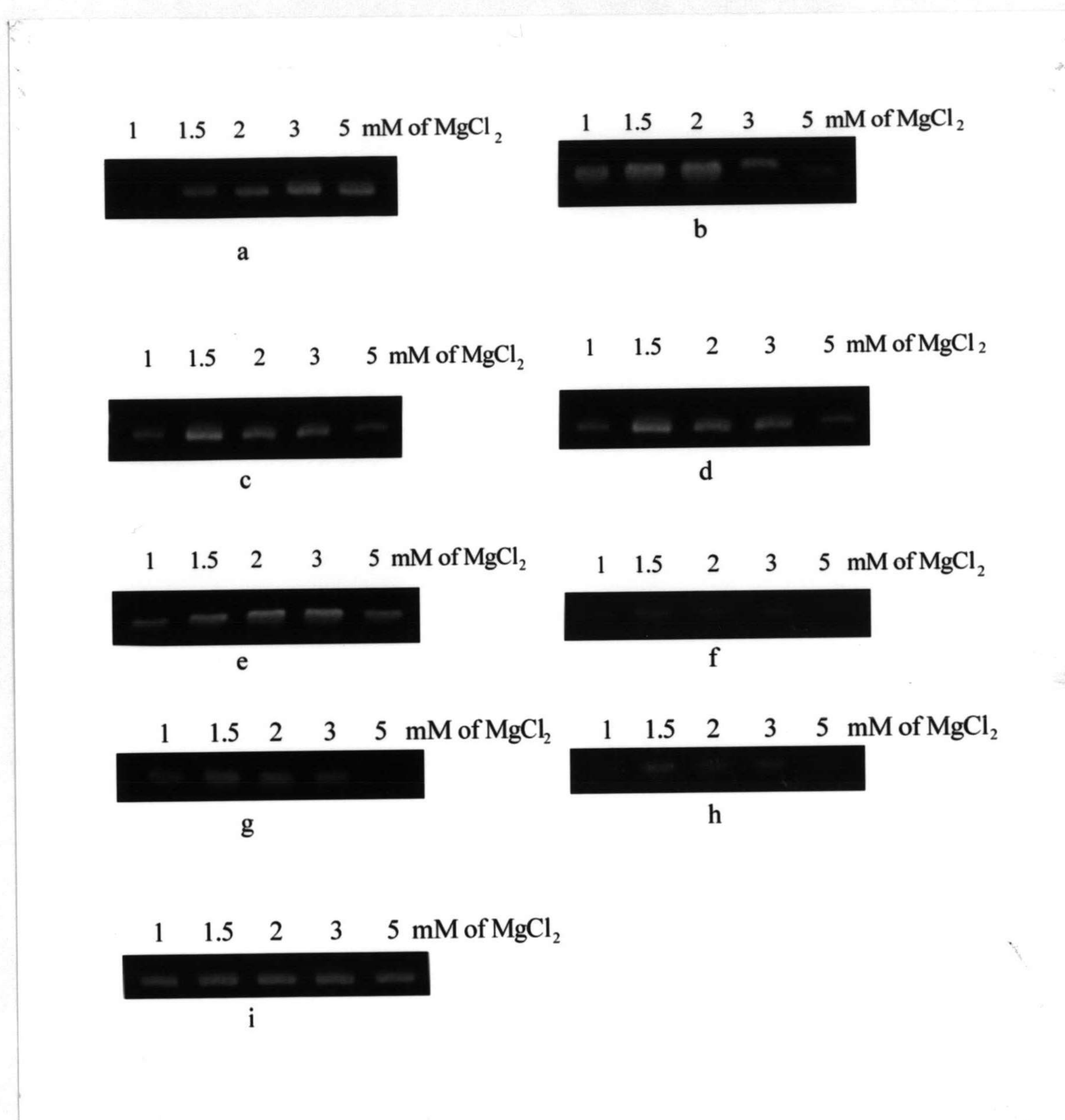

Figure 3.24 Determination of the optimal $MgCl_2$ concentration for PCR amplification by varying concentration of $MgCl_2$ from 1-5 mM.

(a) ALF	(b) SPI
(c) proPO	(d) Crustin
(e) Penaeidin	(f) HSP 90
(g) HSP 70	(h) Lysozyme
(i) β–actin	

Table 3.16 The optimal $MgCl_2$ concentrations for amplification of immune-related gene transcripts

Target mRNA	Primer	$MgCl_2$ (mM)	Product Size (bp)
ALF	5'CGCCAGCAAGATCGTAGGGTTG3' (F)	3	250
	5'AGGCCTATGAGCTGAGCCACTG3' (R)		
SPI	5'TGGCGTGAGTGTCACTTTCCA3' (F)	2	254
	5'AAGTCTTGCCATCACTGCCAC3' (R)		
Crustin	5'TCCCTGGAGGTCAATTCGAGTG3' (F)	1.5	216
	5'AGTCGAACATGCAGGCCTATCC3' (R)		
Penaeidin	5'AGGATATCATCCAGTTCCTG3' (F)	1.5	220
	5'ACCTACATCCTTTCCACAAG3' (R)		
proPO	5'GAGGATATATTTGGCTCCGAAG3' (F)	1.5	419
	5'GGTCGAACGGGAAGCCCATC3' (R)		
HSP 90	5'TCGTCAATACCCAGGCCAA3' (F)	1.5	184
	5'CGACCACAGCATCATCGAAAC3' (R)		
HSP 70	5'GCACCTGCTGTCGGTATTGATC3' (F)	1.5	170
	5'TACAGTGTTGTTGGGGTTCATC3' (R)		
Lysozyme	5'TGGCAGCGATTATGGCAAG3' (F)	1.5	217
	5'GGAACCACGAGACCAGCACTC3' (R)		
β–actin (internal control)	5'GCTTGCTGATCCACATCTGCT3' (F)	1-5	337
	5'ACTACCATCGGCAACGAGA3' (R)		

3.5.3 Determination of cycling parameter

Cycling numbers were important to select the most appropriate number of amplification cycles before reaching a plateau amplification phase. The amplification product showed a sharp DNA band on an agarose gel could then be correctly quantified.

In this experiment, numbers of cycles were determined in a range from 20 to 36 cycles. The number of cycles that gave the highest yield before the product reached a plateau phase was chosen. Four genes (ALF, serine proteinase inhibitor, prophenoloxidase and crustin) showed an approximately contrast yields after 27 cycles (Figure 3.25 a-d). Whereas HSP 90 and HSP 70 showed contrast yields after 30 cycles (Figure 3.25 f-g). Penaeidin, lysozyme and β–actin reached a plateau of amplification after 33 cycles (Figure 3.25 e, h and i). Therefore, 24, 27 and 30 amplification cycles were selected and further investigated for each gene transcript.

3.5.4 Simultaneous amplification of target gene products and an internal control

The semi-quantitative RT-PCR technique is an approach that the target genes and the internal control were simultaneously amplified in the same reaction tube. Therefore, the intensity of the target product and the internal control should be examined when amplification was carried out separately and simultaneously. PCR conditions were further adjusted if necessary.

For simultaneously amplification, competitive effects between target and control primers may be occurred which decreased the intensity of the target and/or the control products when separately amplified. As a result, $MgCl_2$ concentration was further adjusted to eliminate those effects. As can be seen from figure 3.26, $MgCl_2$ concentration of all except crustin and penaedin needed to be increased for 0.5-1.0 mm. Of which, HSP 70 and lysozyme required an additional $MgCl_2$ for 0.5 mM to eliminate effects from primer competition whereas ALF, ProPO, and Penaedin required an additional 1.0 mM $MgCl_2$.

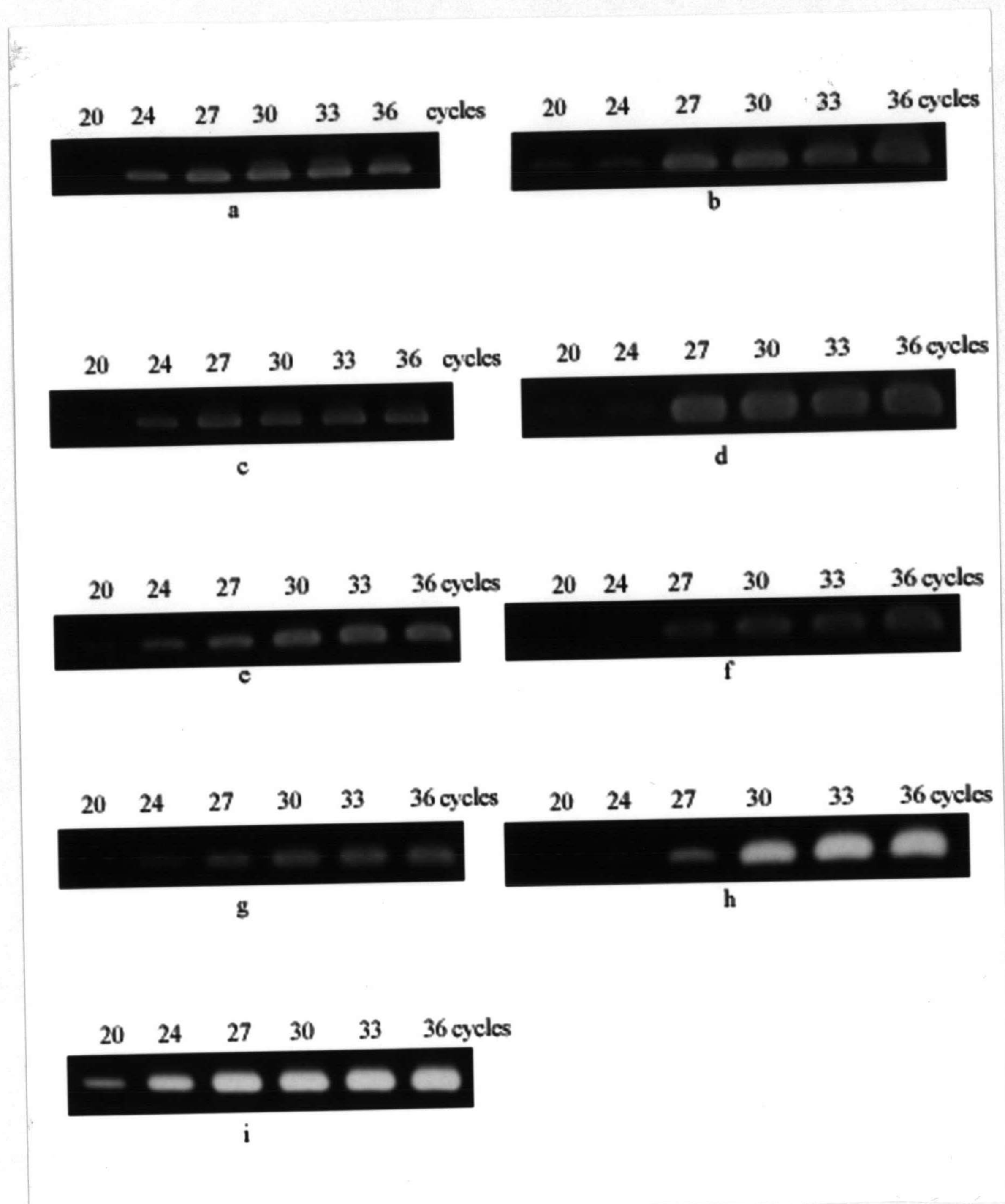

Figure 3.25 Determination of the optimal cycling number of PCR by varying numbers of amplification cycles from 20-36 cycles

(a) ALF
(b) SPI
(c) proPO
(d) Crustin
(e) Penaeidin
(f) HSP 90
(g) HSP 70
(h) Lysozyme
(i) β–actin

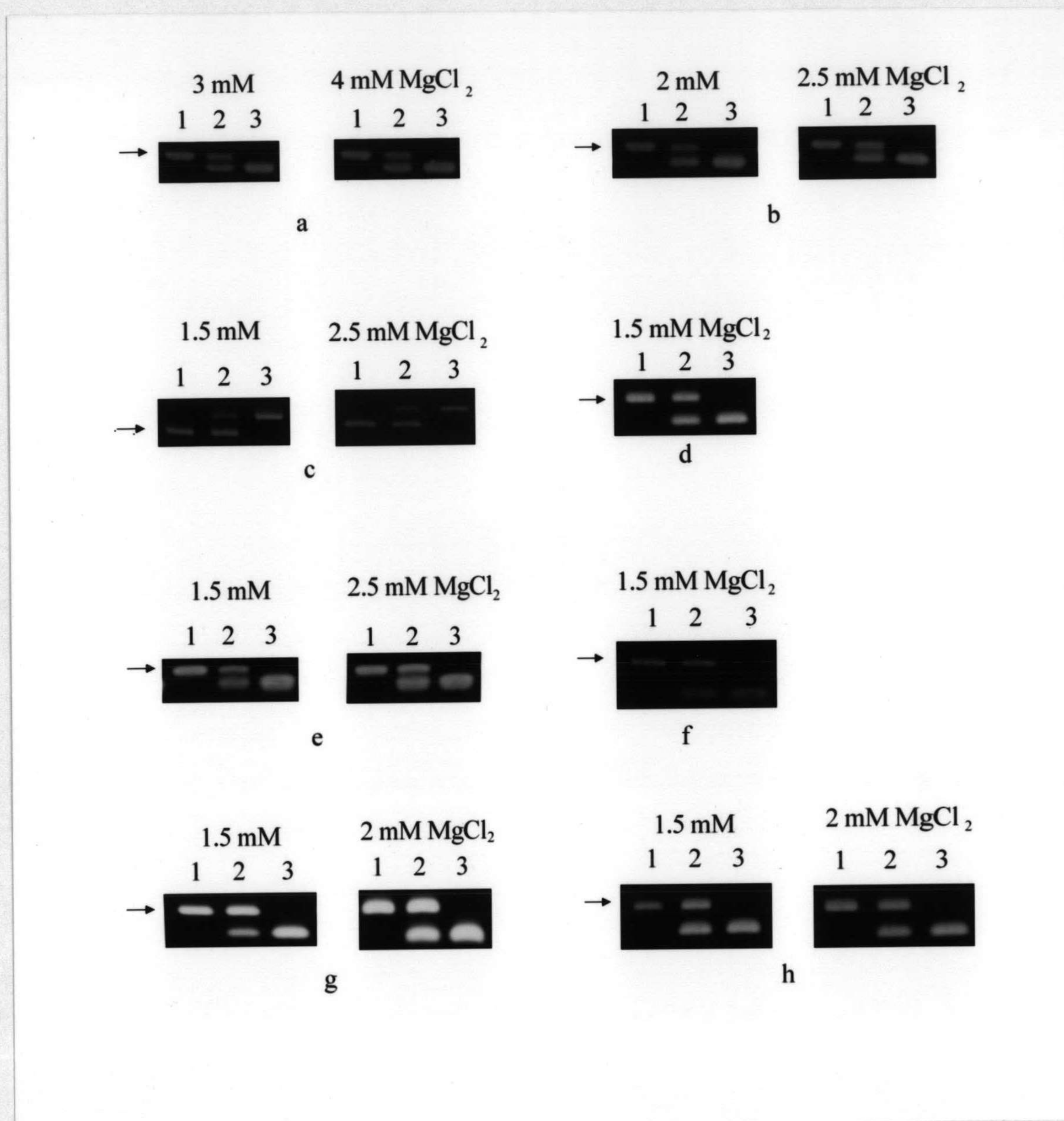

Figure 3.26 Competition between target and internal control primer sets. PCR reactions were simultaneously performed in the same tube. Lane 1:An amplicon from the internal control primer set, lane 3: An amplicon from the target gene primer set, and lane 2: Amplicons from both primer sets. Arrows show PCR product of β-actin.

(a) ALF (b) SPI

(c) proPO (d) Crustin

(e) Penaeidin (f) HSP 90

(g) HSP 70 (h) Lysozyme

The concentrations of $MgCl_2$ used for semi-quantitative RT-PCR amplification of immune-related genes in this study were summarized in Table 3.17.

Table 3.17 The optimal $MgCl_2$ concentration for amplification of immune-related genes

Gene transcript	$MgCl_2$ concentration (mM)
ALF	4.0
SPI	2.5
ProPO	2.5
Penaeidin	2.5
HSP 70	2.0
Lysozyme	2.0
Crustin	1.5
HSP 90	1.5

3.5.5 A time course analysis of mRNA expression level against infection of *V. haeveyi*

Conditions optimized for semi-quantitative PCR were used for a time course analysis of mRNA expression level of unchallenged and *V. harveyi* challenged *P. monodon.*

Sub-adult shrimps were injected with 10^7 cfu/ml of *V. harveyi.* Haemolymph was collected at 0, 6, 12, 24 and 48 hours after injection. At each time point haemolymph from 5 shrimps were pooled. Total RNA were prepared the quantitative experiments were then performed in triplicate using the optimized semi-quantitative RT-PCR technique. The amplification products were run on 1.4% of agarose gel and A ratio of band intensity of the target gene and the control gene was analyzed using Genetools analysis software (Syngene). Results are shown in Table 3.18.

Table 3.18 Expression levels of immune-related gene using a semiquantitative RT-PCR. The experiments were performed in triplicate.

Genes	Relative Expression*					
	0 hour	**3 hours**	**6 hours**	**12 hours**	**24 hours**	**48 hours**
ALF	72.67 (±7.09)	104.00 (±7.81)	116.33 (±23.18)	106.67 (±7.64)	101.33 (±8.145)	90.67 (±6.11)
SPI	85.00 (±5.29)	76.33 (±10.21)	69.33 (±3.51)	81.67 (±13.05)	81.67 (±5.86)	79.00 (±7.94)
ProPO	53.67 (±10.02)	43.00 (±7.55)	46.33 (±11.24)	49.00 (±19.47)	55.67 (±5.03)	51.33 (±13.43)
Crustin	106.67 (±4.73)	41.00 (±9.54)	43.67 (±7.64)	51.33 (±18.45)	72.67 (±10.79)	88.33 (±2.89)
Penaeidin	137.33 (±17.62)	76.00 (±12.12)	57.00 (±11.36)	48.33 (±8.74)	79.33 (±9.29)	107.67 (±23.01)
HSP90	28.33 (±1.15)	36.00 (±3.00)	40.00 (±9.17)	42.33 (±2.08)	36.00 (±6.25)	30.00 (±2.00)
HSP70	93.67 (±14.19)	92.33 (±12.10)	87.33 (±10.02)	92.00 (±10.82)	89.67 (±3.79)	88.33 (±3.51)
Lysozyme	38.33 (±15.01)	46.00 (±13.89)	60.00 (±12.17)	65.33 (±6.81)	66.33 (±9.61)	73.33 (±4.93)

* The expression was determined as the signal ratio of the interest gene: β- actin while the expression of β- actin was normalized to 100.

After injection with *V. harveyi*; expression levels of ALF (Figures 3.27a, 3.28), HSP 90 (Figures 3.27b, 3.29) and lysozyme (Figures 3.27c, 3.30) were significantly increased ($p<0.05$, Data are shown in appendix B). Significant expression levels of ALF was observed since 3 hours after injection and was not significantly lowered until 48 hours after injection (72.67 for normal shrimps and 90.67-116.33 for challenged shrimps, $p<0.05$). The highest level was 1.6 times above that of normal shrimps at 6 hours after injection. Likewise, expression of HSP 90 was increased significantly after challenged with *V. harveyi* ($p<0.05$). The trend of expression of HSP 90 was similar as that of ALF and the highest level of HSP 90 expression was observed at 12 hours after injection at 1.5 times above the normal condition. Significant expression level of lysozyme was observed since 6 hours after injection ($p<0.05$) and still increased during the study period. The highest expression level of lysozyme was observed at 48 hours after injection at 1.9 times above the normal condition.

Injection of *V. harveyi* caused a significant decrease in the expression of crustin and penaeidin ($p<0.05$, Data are shown in appendix B). The expression of crustin was significantly lower than that of the normal since 3 hours after injection of *V. harveyi* until the end of assay at 48 hours (Figures 3.31a, 3.32). The lowest expression level of crustin in *P. monodon* was 0.38 times below the normal shrimp at 3 hours after injection. Although expression of penaeidin was significantly decreased after 3 hours of challenged ($p<0.05$) (Figures 3.31b, 3.33), the lowest expression was observed at 12 hours after *V. harveyi* injection (0.35 times that of the normal shrimp). After reaching the lowest level, the expression of crustin and penaeidin were increased significantly after 24 hours ($p<0.05$) and 48 hour of injection ($p<0.05$), respectively, but the expression level of these time points of both proteins were still lower than that of normal shrimps. Constitutive expression of serine proteinase inhibitor, prophenoloxidase, HSP 70 and lysozyme was observed (Figures 3.33-3.37). The expression levels of these genes were not significantly different to the normal condition at all time points ($p>0.05$, Data are shown in appendix B).

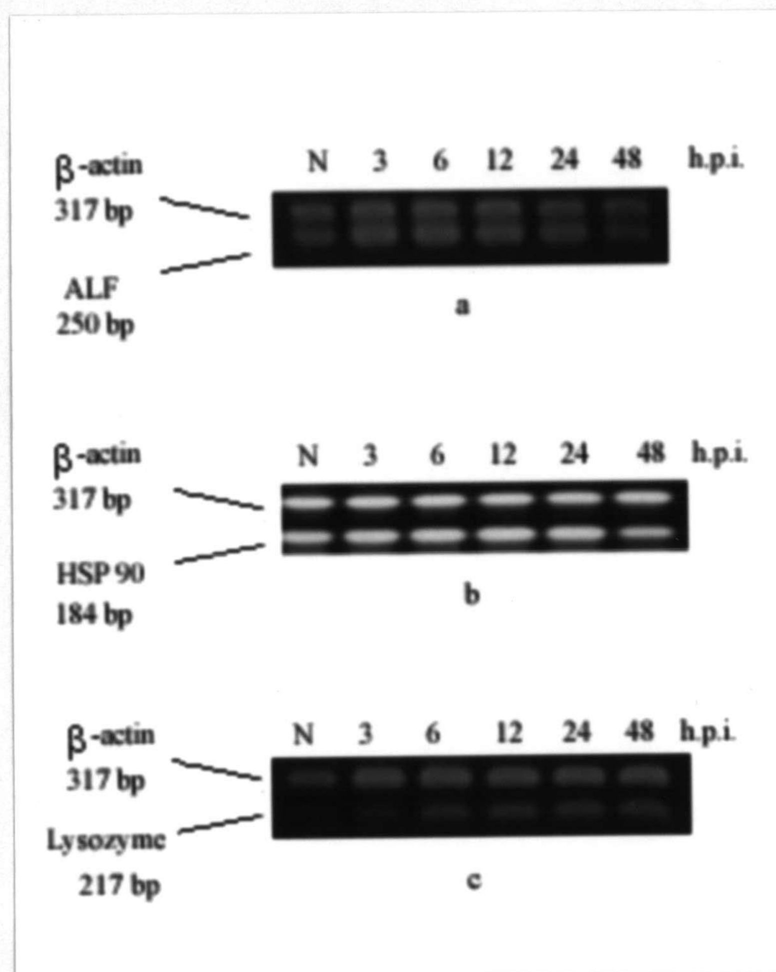

Figure 3.27 Analysis of expression levels of ALF (a), HSP 90 (b) and lysozyme (c) transcripts after injection with *V. harveyi* for 3, 6, 12, 24, 48 hrs. N; Normal shrimp.

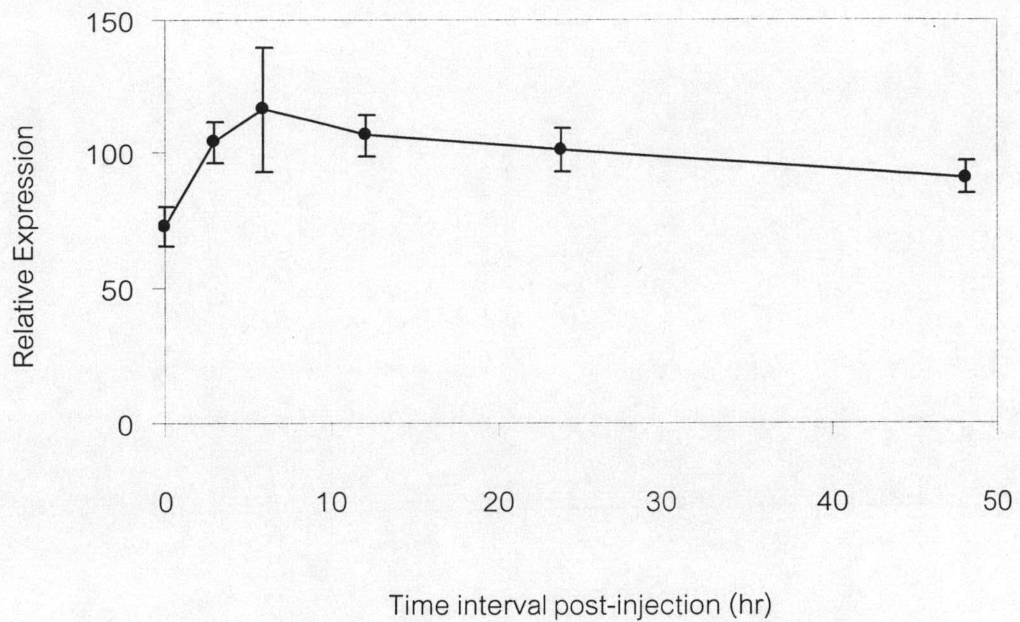

Figure 3.28 Relative expression levels of ALF at different time of intervals after injected with *V. harveyi*.

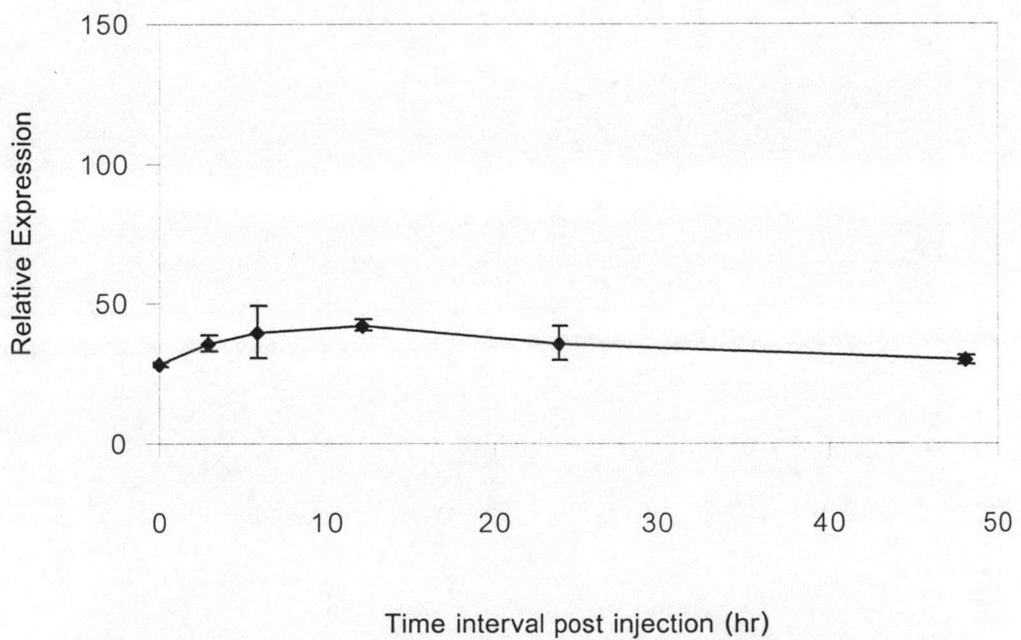

Figure 3.29 Relative expression levels of HSP90 at different time of intervals after injected with *V. harveyi*.

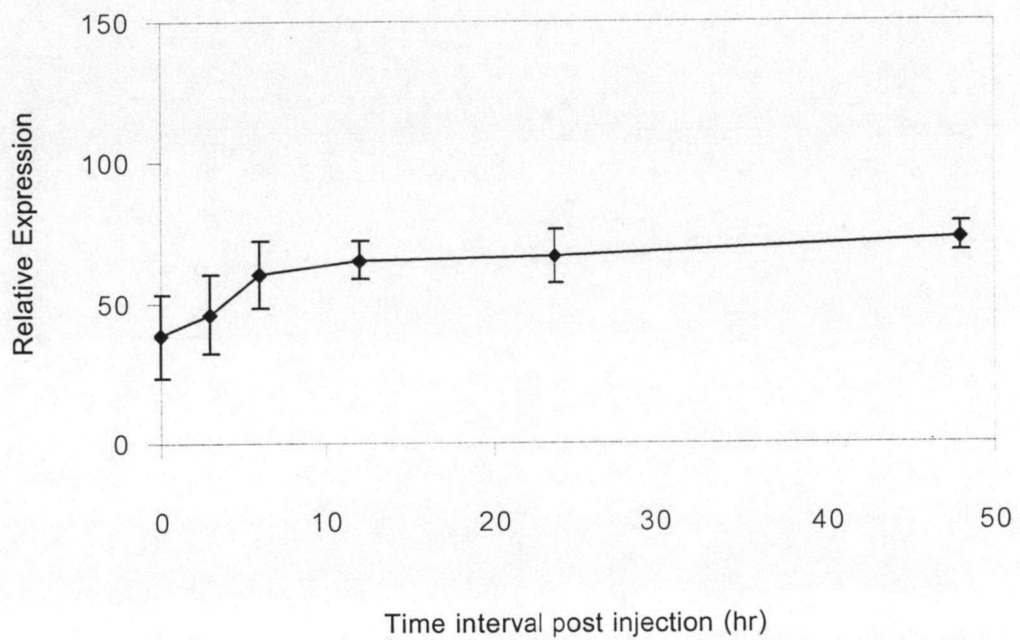

Figure 3.30 Relative expression levels of lysozyme at different time of intervals after injected with *V. harveyi*.

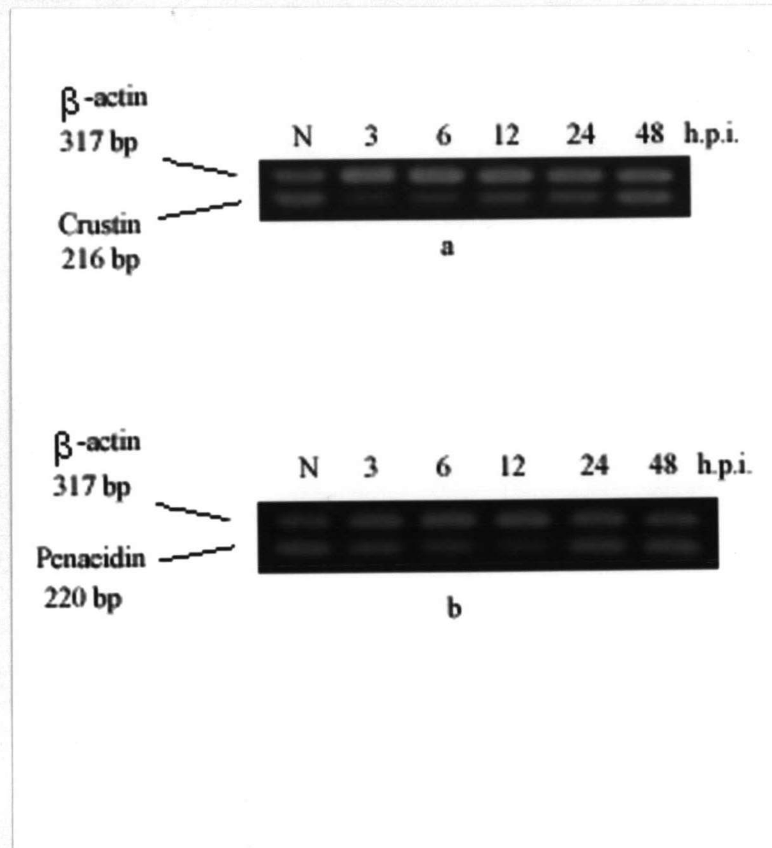

Figure 3.31 Analysis of expression level of crustin (a) and penaeidin (b) transcripts after injection with *V. harveyi* for 3, 6, 12, 24, 48 hrs. N; Normal shrimp.

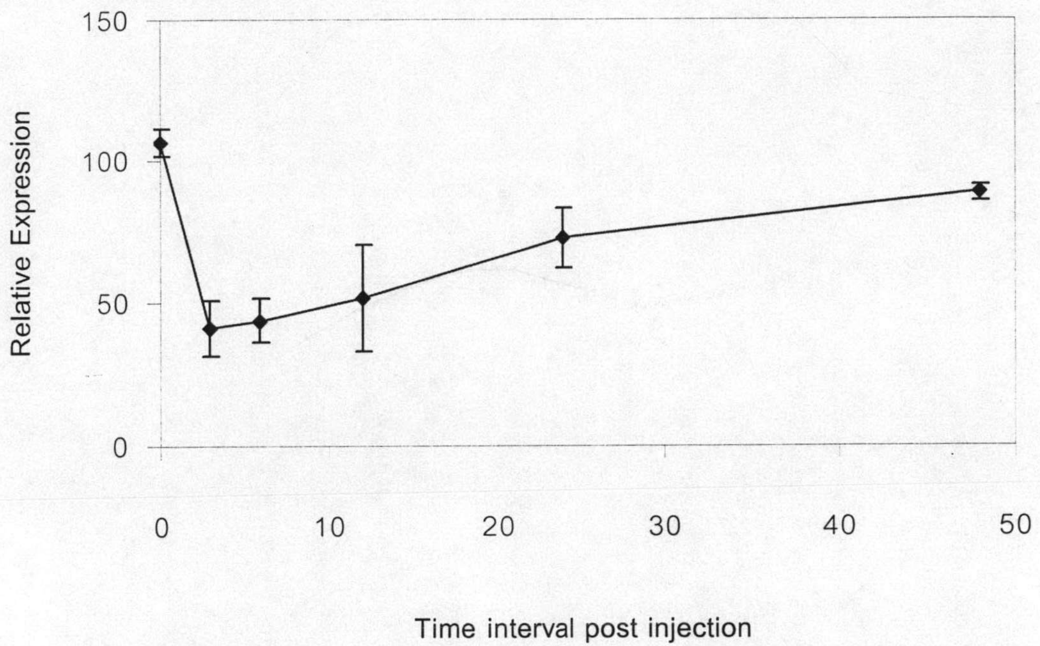

Figure 3.32 Relative expression levels of crustin at different time of intervals after injected with *V. harveyi.*

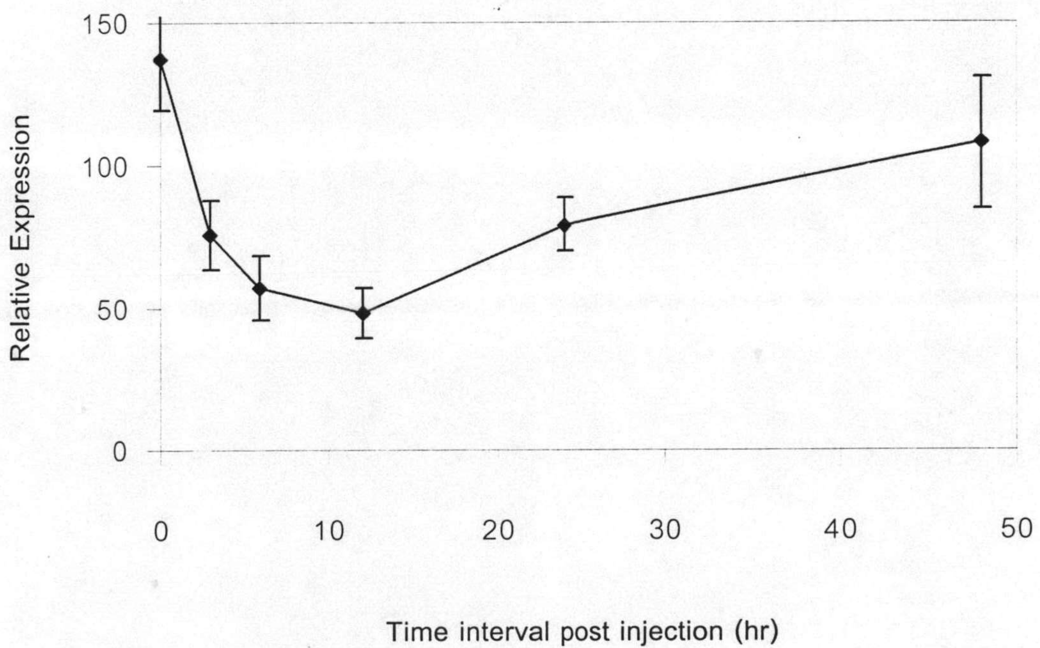

Figure 3.33 Relative expression levels of Penaeidin at different time of intervals after injected with *V. harveyi.*

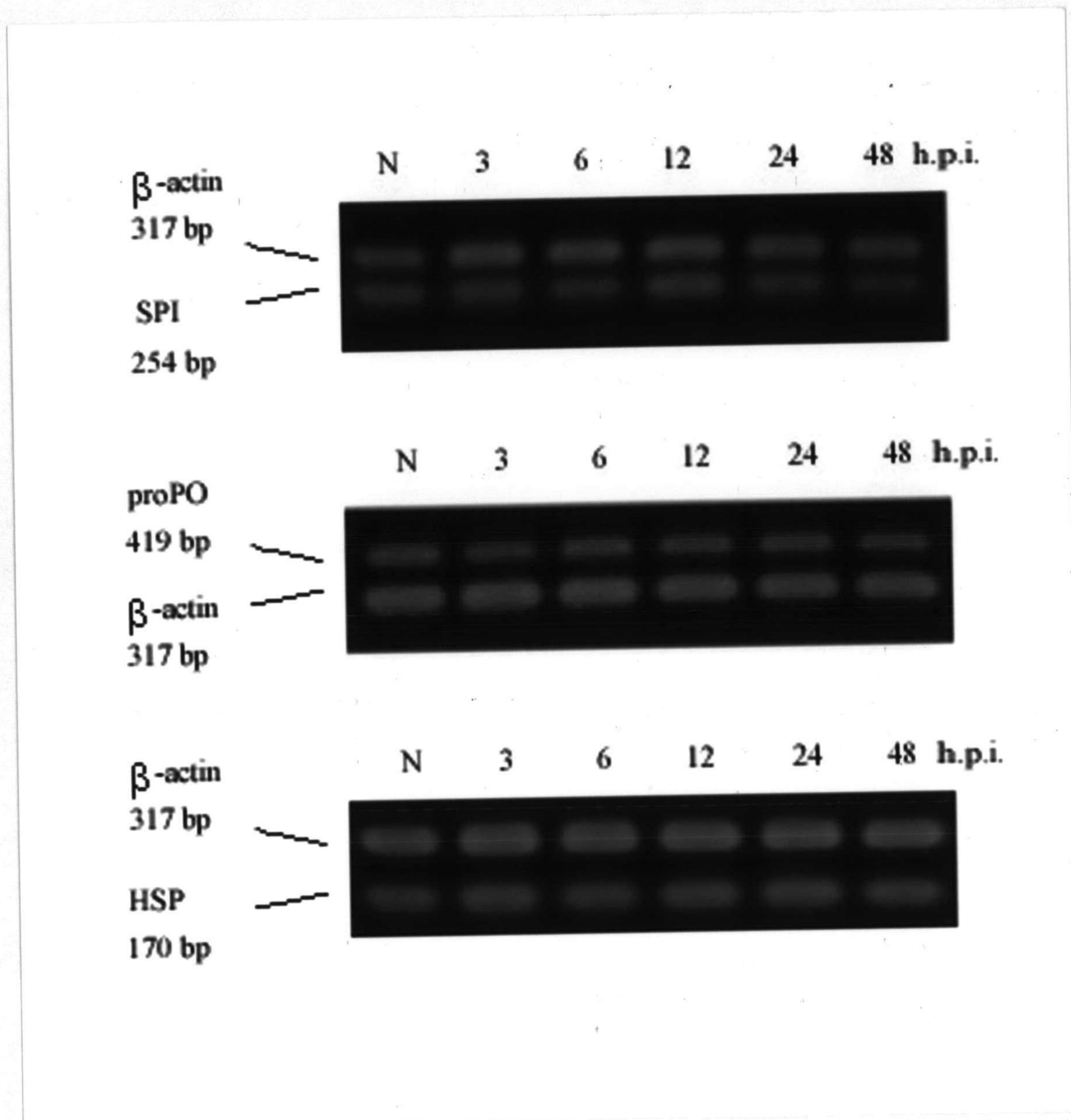

Figure 3.34 Analysis of expression level of srine proteinase inhibitor (a), prophenoloxidase (b) and HSP 70 (c) transcripts after injection with *V. harveyi* for 3, 6, 12, 24, 48 hrs. N; Normal shrimp.

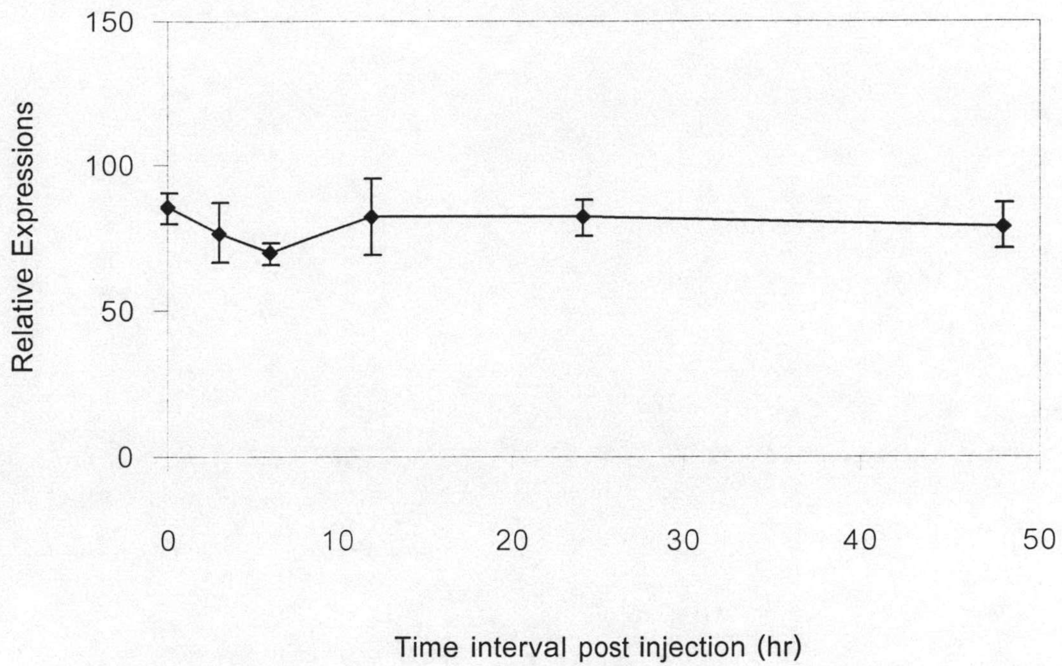

Figure 3.35 Relative expression levels of SPI at different time of intervals after injected with *V. harveyi*.

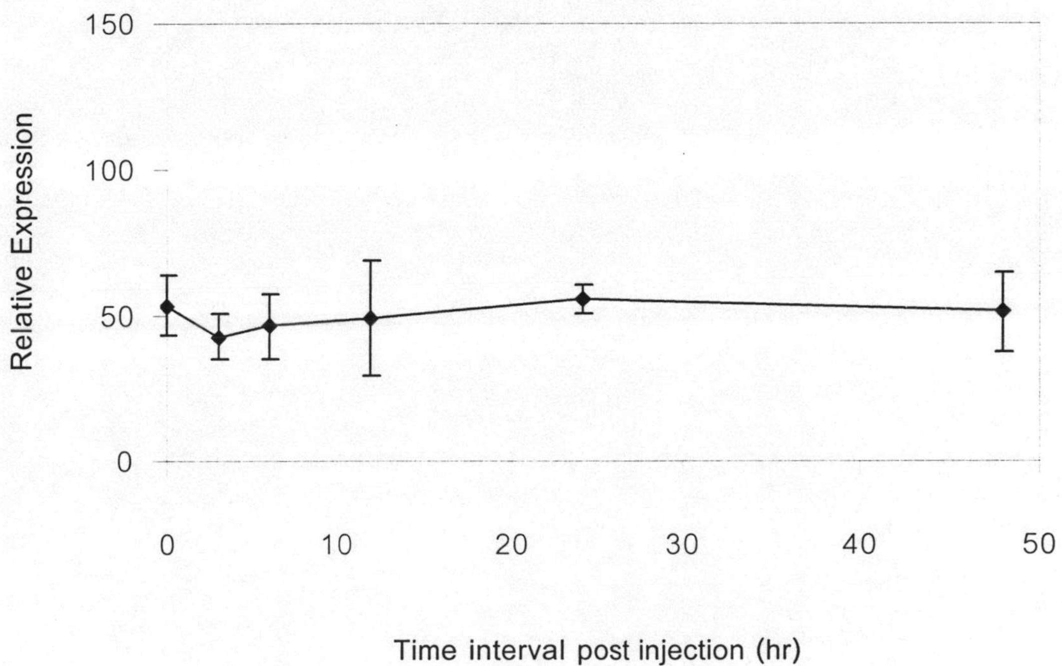

Figure 3.36 Relative expression levels of proPO at different time of intervals after injected with *V. harveyi*.

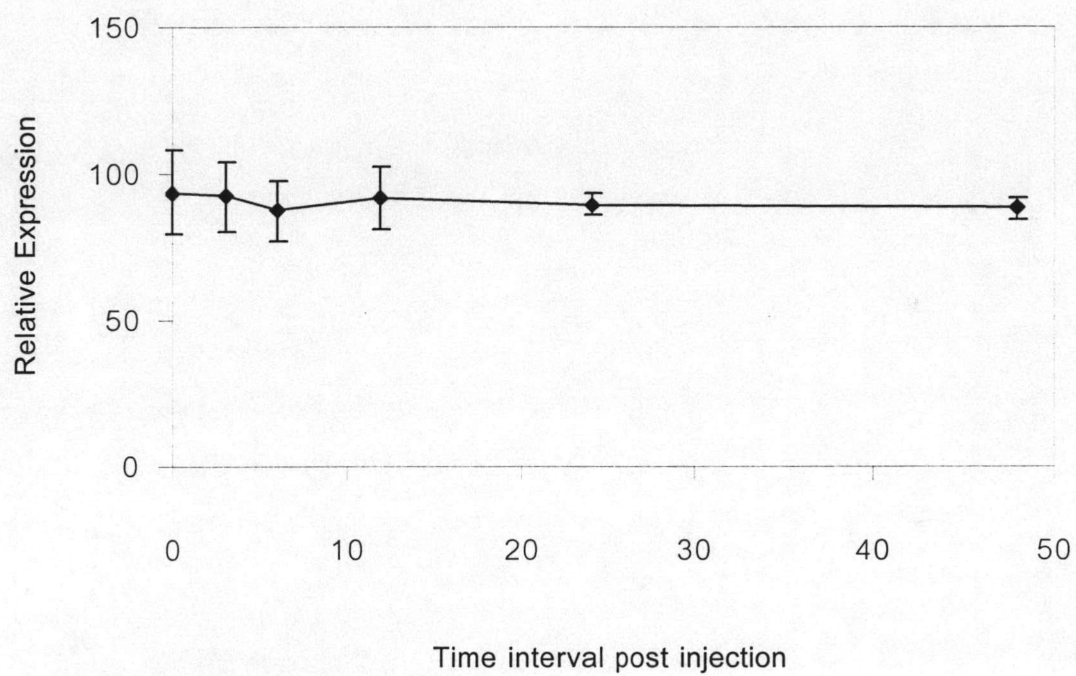

Figure 3.37 Expression levels of HSP70 at different time intervals after injected with *V. harveyi*.