# CHAPTER III

# RESULTS

### 3.1 RT-PCR of functionally important genes

Total RNA from ovaries and testes revealed predominant discrete bands along with smeared high molecular weight RNA (Figure 3.1).The ratios of extracted RNA were 1.7 - 2.0. The first strand cDNA synthesized from those total RNA covered the large products indicating the acceptable quality of the synthesized first strand cDNA (Figure 3.2)

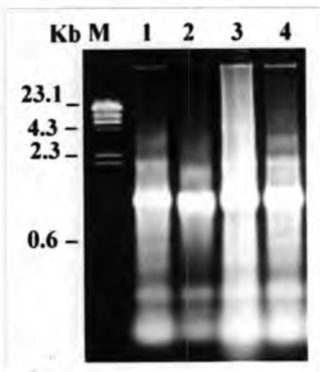



**Figure 3.1** A 1.0% ethidium bromide-stained agarose gel showing the quality of total RNA extracted from ovaries of *P. monodon*. Lane M = λ-*Hin*d III. Lane 1 - 4 = Total RNA individually extracted from ovaries of each *P. monodon* broodstock.

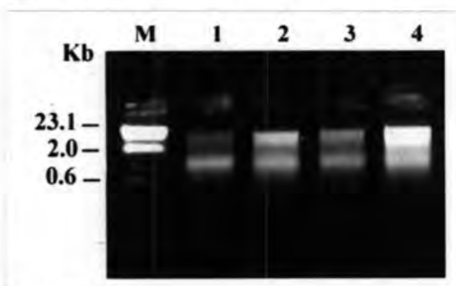



**Figure 3.2** A 1.0% ethidium bromide-stained agarose gel showing the synthesized first strand cDNA from ovaries of *P. monodon* and Lane M = λ-*Hin*d III. Lane 1 - 4 = the first strand cDNA from ovaries of each *P. monodon* broodstock.

A total of 158 primer pairs were designed from EST libraries of ovaries (71 primers pairs) and testes (2 primer pairs) of normal shrimp and hemocytes of temperature-stressed *P. monodon*, (85 primer pairs). RT-PCR was carried out using an identical amplification condition for all primer pairs.

Initially, the first strand cDNA synthesized from ovaries and testes of 2 individuals of each sex were subjected to RT-PCR and electrophoretically analyzed. A total of 111 primer pairs generated the positive amplification products. Of these, 9 gene homologues were expressed only in ovaries. This included homologues of *female sterile, ATP/GTP binding protein, adipose differentiation related protein, broad complex Z4 isoform, ovarian lipoprotein receptor, carbonic anhydrate, aminopeptidase, Wolf hirschorn syndromè candidate 1 protein*, and *proactivator polypeptide precursor* (Figures 3.3 – 3.11; Table 3.1)

Sixty-four gene homologues were preferentially expressed in ovaries than testes of *P. monodon* broodstock. Examples of these genes were *Y-box protein p-50, 3-oxoacid CoA transferase, ferritin signal crayfish, small androgen receptor interacting protein, NADP- dependent leukotriene B4 12 hydroxy dehydrogenase, dolichyl diphosphooligocharide protein glycotransferase, asparaginyl tRNA synthetase, aspartate amino transferase*, and *nuclear autoantigenic sperm protein* (Figures 3.14 – 3.20; Table 3.2).

Nevertheless, 38 gene homologue (e.g. *dendritic cell protein* and *calponin I*) did not revealed differential expression between ovaries and testes of female and male *P. monodon* broodstock (Figure 3.21; Table 3.3). In addition, 5 primer pairs generated non-specific amplification products (Table 3.4) and 42 primer pair did not provide the positive amplification product (Table.3.5).

Seventy-three gene homologues that exhibited specific or preferential expression in ovaries of *P. monodon* were further screened with a larger sample size ($N$ = 4 for each sex). Results from the secondary screening agree with those of the primary screening.

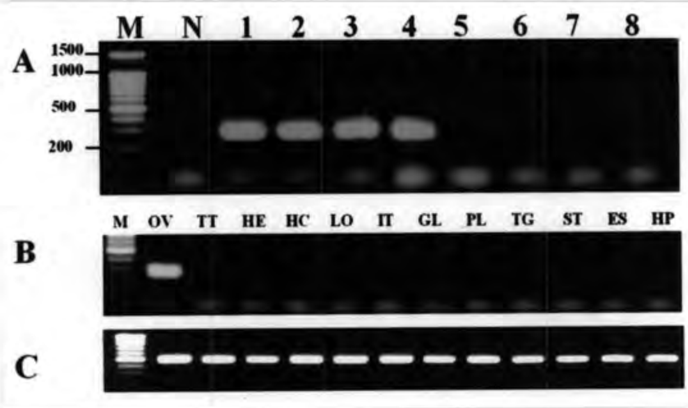



**Figure 3.3** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *female sterile* using the first strand cDNA of ovaries (lanes 1 – 4, A) and testes (lane 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. The positive product of *EF-1α* was successfully amplified from the same template in A and B (C). Lanes M and N are a 100 bp DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

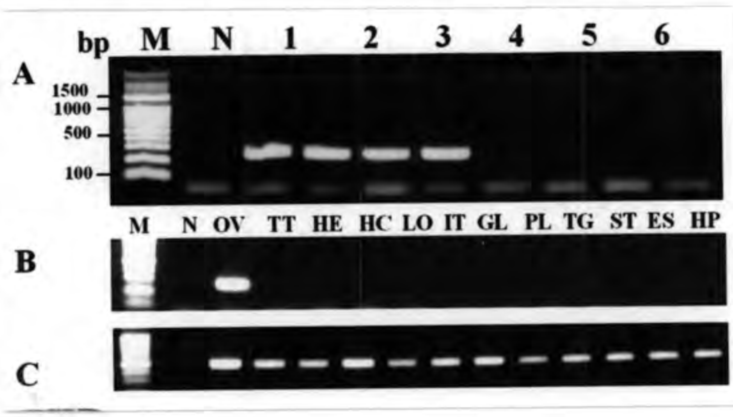



**Figure 3.4** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *ovarian lipoprotein receptor* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues of broodstock-sized *P. monodon*. The positive product of *EF-1α* was successfully amplified from the same template in A and B (C).Lanes M and N are a 100 bp DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.


**Figure 3.5** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *ATP/GTP binding protein* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. The positive product of *EF-1α* was successfully amplified from the same template in A and B (C).Lanes M and N are a 100 bp DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.


**Figure 3.6** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *adipose differentiation related protein* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. The positive product of *EF-1α* was successfully amplified from the same template in A and B (C).Lanes M and N are a 100 bp DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.


**Figure 3.7** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *Wolf hirschorn syndrome candidate 1 protein* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. The positive product of *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.


**Figure 3.8** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *broad complex Z4 isoform* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. The positive product of *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp DNA marker and the the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.


**Figure 3.9** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *aminopeptidase* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. The positive product of *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp DNA marker and the the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas..

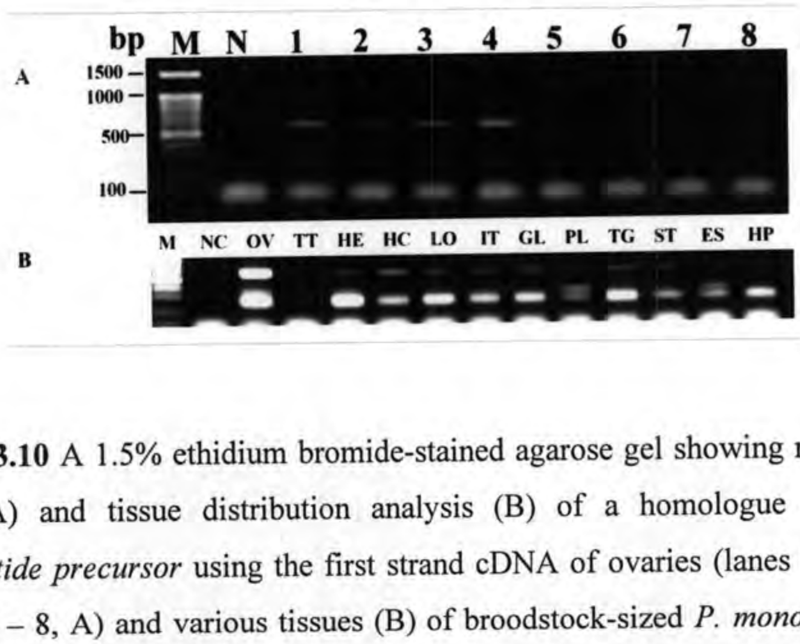



**Figure 3.10** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *proactivator polypeptide precursor* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B Lanes M and N are a 100 bp DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

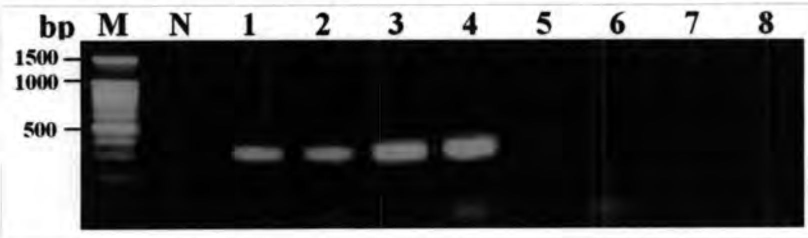



**Figure 3.11** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR of a homologue of *carbonic anhydrate* using the first strand cDNA of ovaries (lanes 1 - 4) and testes (lanes 5 - 8) of broodstock-sized *P. monodon*. Lanes M and N were a 100 bp DNA marker and the negative control, respectively

**Table 3.1** Gene specifically expressed in ovaries but not testes of *P. monodon*

| Gene homologues | Expected size.(bp.) | Tissue |
|---|---|---|
| 1. *Female sterile* | 296 | O |
| 2. *Ovarian lipoprotein receptor* | 354 | O |
| 3. *ATP/GTP binding protein* | 360 | O, HE, HC, LO, IT, GL, PL, TG, ST |
| 4. *Wolf hirschorn syndrome candidate 1 protein* | 208 | O,IT,LO,HE |
| 5. *Adipose differentiation related protein* | 267 | O, HE, HC, LO, TG, ST |
| 6. *Aminopeptidase* | 174 | O, HE, HC,LO, IT, GL,ST |
| 7. *Proactivator polypeptide precursor* | 266 | O, HE, HC, LO, IT, GL, PL, TG, ST, ES, HP |
| 8. *Broad complex Z4 isoform* | 192 | O, HC, LO, IT, GL, PL, TG, ST, ES, HP, |
| 9. *Carbonic anhydrase* | 332 | Not determined |

Tissue distribution analysis was carried out to examine the expression of 26 gene homologues in ovaries, heart, hemocytes, lymphoid organs, intestine, gills, pleopods, thoracic ganglion, stomach, eyestalk and hepatopancreas of a single female broodstock and testes of a single male broodstock of *P. monodon*.

For gene homologues expressed only in ovaries but not testes, female sterile and ovarian lipoprotein receptor were specifically expressed in ovaries but not other examined tissues (Figures 3.3 and 3.4; Table 3.1) suggesting that these genes play an important role in ovaries of *P. monodon*.

Four gene homologues of this group; *ATP/GTP binding protein* (Figure 3.5), *adipose differentiation related protein* (Figure 3.6), *Wolf hirschorn syndrome candidate 1 protein* (Fig. 3.7) and *aminopeptidase* (Figure 3.9) were abundantly expressed in ovaries whereas low expression levels were observed in other tissues (Table 3.6). *Broad complex Z4 isoform* exhibited the highest expression level in ovaries and showed a moderate expression level in hemocytes. Low expression levels of this gene were found in other tissues (Figure 3.8). *Proactivator polypeptide precursor* was highly expressed in ovaries but moderately expressed in most of the remaining tissues (Figure 3.10).

Tissue distribution analysis further confirmed the preferential expression of 16 genes in ovaries than testes of *P. monodon* broodstock. *Y-box protein* (Figure 3.14), *NADP dependent leucotrience B4 12 hydroxy dehydrogenase* (Figure 3.15), *dolichyl diphosphooligocharide protein glycotransferase* (Figure 3.16) and *nuclear autoantigenic sperm protein* (Figure 3.17) were highly expressed in ovaries than other tissues and lower in testes and intestine but not in other tissue.

On the other hand, expression of *asparaginyl tRNA synthetase* (Figure 3.18) in stomach was greater than that in ovaries, hemocytes and other tissues. Likewise, a homologue of *3-oxoacid CoA transferase*, was expressed in stomach and ovaries are comparable and higher than other tissues. (Figure 3.19). *O-methyl transferase* gene was highly expression in hemocyte and hepatopancrease. Moderate expression was observed in ovaries and low expression was found in the remaining tissues (Figure 3.20)

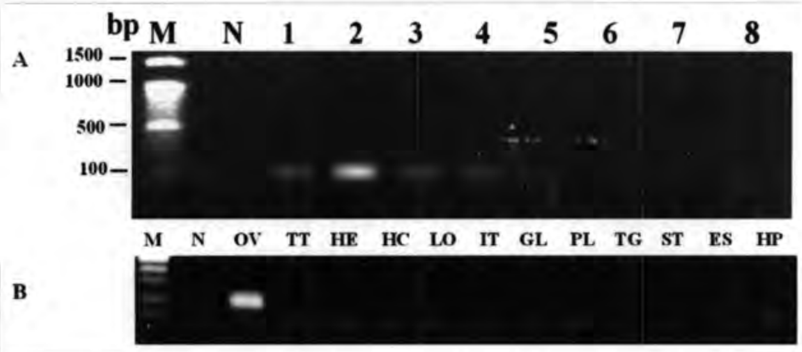



**Figure 3.12** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *zonadhesin isoform 4* using the first strand cDNA of ovaries (lanes 1 - 4) and testes (lanes 5 - 8) and various tissues of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp. DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

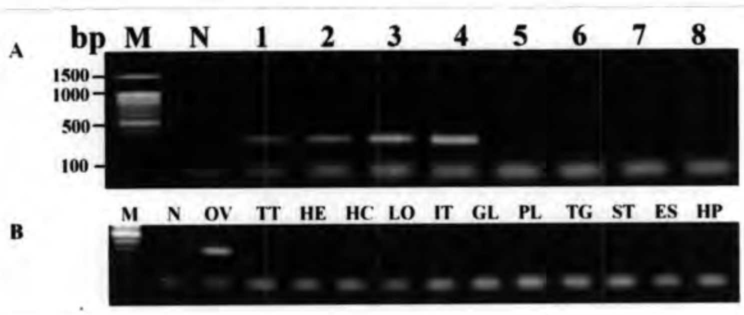



**Figure 3.13** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *TATA binding protein associated factor9* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp. DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.


**Figure 3.14** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *Y-box protein* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp. DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.


**Figure 3.15** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *NADP dependent leucotrience B4 12 hydroxy dehydrogenase* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp. DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

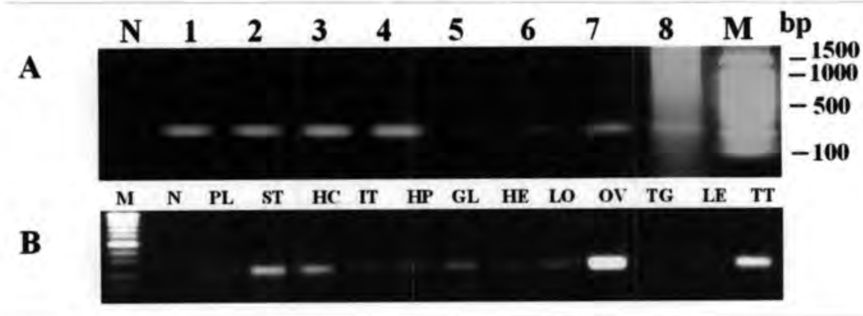



**Figure 3.16** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *dolichyl diphosphooligocharide protein glycotransferase* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp. DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

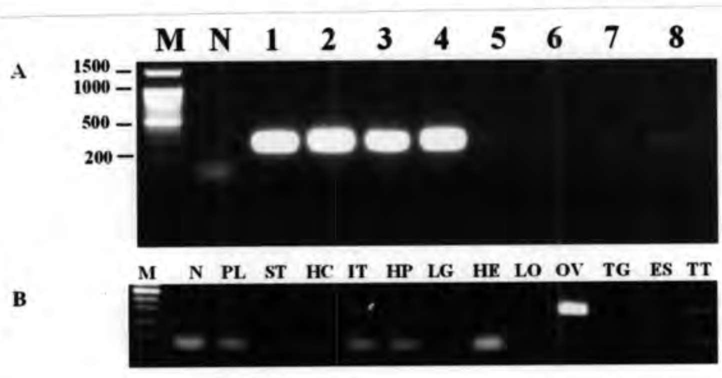



**Figure 3.17** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *nuclear autoantigenic sperm protein* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp. DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

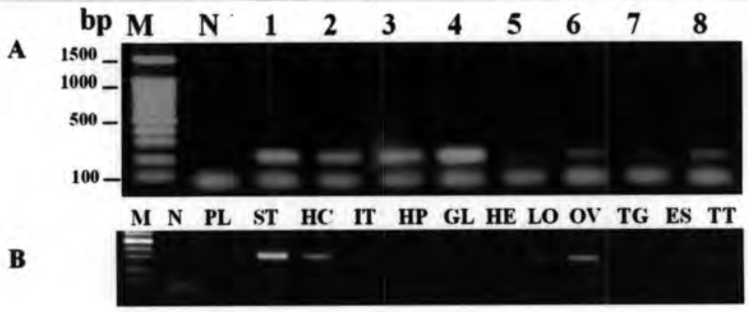



**Figure 3.18** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *asparaginyl tRNA synthetase* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp. DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

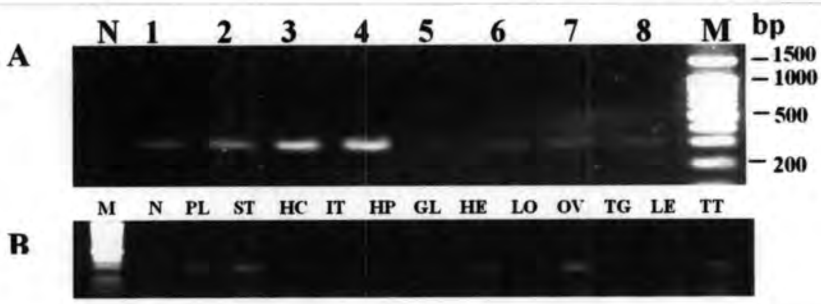



**Figure 3.19** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *3-oxoacid CoA transferase* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp. DNA marker and the negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

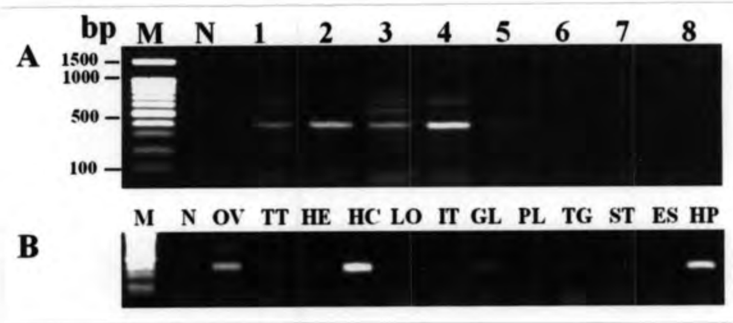



**Figure 3.20** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR (A) and tissue distribution analysis (B) of a homologue of *O-methyl transferase* using the first strand cDNA of ovaries (lanes 1 – 4, A), testes (lanes 5 – 8, A) and various tissues (B) of broodstock-sized *P. monodon*. *EF-1α* was successfully amplified from the same template in A and B. Lanes M and N are a 100 bp DNA marker and negative control, respectively. OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

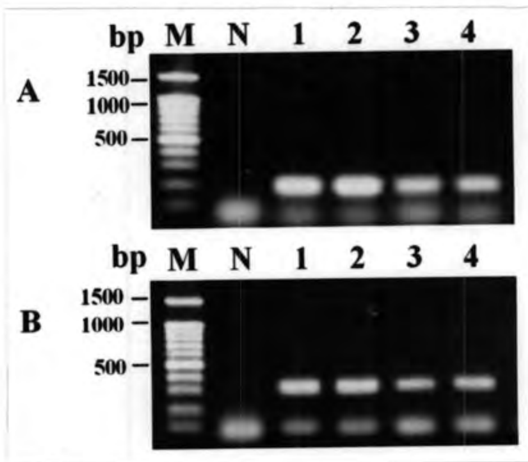



**Figure 3.21** A 1.5% ethidium bromide-stained agarose gel showing results from RT-PCR of a homologue of *dendritic cell protein* (A) and *calponin I* (B) using the first strand cDNA of ovaries (lanes 1 – 2), testes (lanes 3 – 4) of broodstock-sized *P. monodon*. Lanes M and N are a 100 bp DNA marker and the negative control, respectively.

**Table 3.2 Gene homologues exhibiting preferential expression patterns in ovaries than testes of *P. monodon***

| Gene homologues | Expected size.(bp.) | Tissue distribution |
|---|---|---|
| 1. *NADH dehydrogenase subunit 5* | 250 | Not determined |
| 2. *Phenylalanine-tRNA synthesis-B-subunit* | 221 | Not determined |
| 3. *3-oxoacid CoA transferase* | 303 | Not determined |
| 4. *Prefold in subunit 2* | 208 | Not determined |
| 5. *CalcineurinB* | 319 | Not determined |
| 6. *Phosphatidyl inosital 4 kinase* | 335 | Not determined |
| 7. *Carboxyl esterase precursor* | 282 | Not determined |
| 8. *US small nuclear ribonucleoprotein* | 277 | Not determined |
| 9. *Stromal membrane associated protein* | 306 | Not determined |
| 10. *Phosphopyruvate hydratase* | 233 | Not determined |
| 11. *Ferritin* | 280 | Not determined |
| 12. *Postsynaptic density protein (citron)* | 212 | Not determined |
| 13. *Kin protein* | 188 | Not determined |
| 14. *USO1* | 314 | Not determined |
| 15. *Small androgen receptor interacting protein* | 345 | Not determined |
| 16. *Proteosome (proteosome subunit alpha type3)* | 283 | Not determined |
| 17. *Signal recognition particle 72 kDa* | 291 | Not determined |
| 18. *RUVB like protein2* | 257 | OV, TT, HE, HC, LO, IT, GL, PL, TG, ST, ES, HP |
| 19. *Dynectin 4* | 300 | Not determined |
| 20. *Contractile ring component anillin* | 199 | Not determined |
| 21. *Chromobox protein* | 330 | OV, TT, HE, HC, LO, IT, GL, PL, TG, ST |
| 22. *Rho protein* | 238 | Not determined |
| 23. *Nm2 protein* | 365 | Not determined |
| 24. *Cardiomyopathy associated 4 sterile muscle* | 294 | Not determined |
| 25. *Cytoplasmic actin depolymerizing factor* | 253 | Not determined |

**Table 3.2 (cont.)**

| Gene homologues | Expected size (bp.) | Tissue distribution |
|---|---|---|
| 26. *Semaphorin 2A precursor* | 256 | Not determined |
| 27. *Heterogeneous nuclear RNA protein clone* | 350 | Not determined |
| 28. *Poliferating cell nucleolar antigen p120* | 235 | Not determined |
| 29. *Thiolase* | 231 | Not determined |
| 30. *Profilin* | 229 | Not determined |
| 31. *Ribophorin I* | 190 | Not determined |
| 32. *AgCP13148* | 212 | Not determined |
| 33. *Glycogen phospholipase* | 188 | Not determined |
| 34. *Mapre 1 protein* | 205 | Not determined |
| 35. *Calcium independent phospholipase A2isoform1* | 164 | Not determined |
| 36. *Zeta 1 cop* | 191 | Not determined |
| 37. *Calcium regulated heat stable protein* | 263 | Not determined |
| 38. *Dynein heavy chain* | 192 | Not determined |
| 39. *L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain* | 221 | Not determined |
| 40. *Dolichyl diphosphooligocharide protein glycotransferase* | 233 | OV, TT, HE, HC, LO, IT, GL, PL, TG, ST, ES, HP |
| 41. *Asparaginyl tRNA synthetase* | 330 | OV, TT, HC, IT, ST, ES, LO, PL, GL |
| 42. *Myosin regulatory light chain polypeptide 91* | 300 | Not determined |
| 43. *ESO 3 protein* | 349 | Not determined |
| 44. *Rasputin CG 9412* | 248 | Not determined |
| 45. *Hepatocarcinogenesis related transcription factor (x box protein)*18. *RuvB like protein2* | 185 | Not determined |
| 46. *Carnitine palmitoyl transferase II* 19 | 334 | Not determined |
| 47. *Y-box protein p-50*20 | 433 | OV, TT, HE, HC, IT, GL, PL, LO, TG, ST, ES, HP |
| 48. *RAB protein 10 CG 17060-PA* 21. | 177 | OV, TT, HE, HC, LO, IT, GL, PL, TG, ST, ES |

**Table 3.2** (cont.)

| Gene homologues | Expected size.(bp.) | Tissue distribution |
|---|---|---|
| 49. *Tissue specific transplantation antigen p35B like protein* | 352 | Not determined |
| 50. *Fus prove protein* | 353 | Not determined |
| 51. *Methyl cytosine* | 380 | Not determined |
| 52. *CG-1681-PA* | 243 | OV, TT, HE, HC, LO, IT, GL, PL, TG, ST, ES, HP |
| 53. *Nuclear autoantigenic sperm protein* | 301 | OV, TT, IT |
| 54. *Phosphatidyl serine receptor short form* | 300 | OV, TT, HE, HC, LO, IT, GL, PL, TG, ST, ES |
| 55. *Singed protein* | 214 | Not determined |
| 56. *Survival motor neuron* | 245 | Not determined |
| 57. *Hypothetical protein XP 207715 cyclin nucleotide( cyclic nucleotide gated channel beta subunit 1)* | 270 | Not determined |
| 58. *Solute carrier family 25, member 5;2F1:adenine nucleptide transiocation2 (Fibroblast)* | 239 | Not determined |
| 59. *Ferrochelatase* | 215 | Not determined |
| 60. *NADP dependent leukotriene B4 12 hydroxy dehydrogenase* | 230 | OV, TT, HE, HC, LO, IT, PL, TG, ST, ES, HP |
| 61. *Zonadhesin isoform 4* | 177 | OV, TT, HE, HC, LO |
| 62. *TAF15 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 68kDa. isoform CRA_c* | 320 | Not determined |
| 63. *Immunophilin FKBP 52* | 311 | Not determined |
| 64. *Aspartate aminotransferase* | 334 | OV,TT,HE,HC,IT,PL,ES,HP |

OV = ovaries, TT = testes, HE = heart, HC = hemocytes, LO = lymphoid organs, IT = intestine, GL = gill, PL = pleopods, TG = thoracic ganglion, ST = stomach, ES = eyestalk and HP = hepatopancreas.

**Table 3.3** Gene homologues exhibiting identical expression levels in ovaries and testes of *P. monodon* broodstock

| Gene homologues | Expected size.(bp.) |
|---|---|
| 1. *Guanine nucleotide binding protein* | 180 |
| 2. *Thyroid hormone receptor associated protein complex 230 kDa component* | 312 |
| 3. *Myelodysplasia/Myeloid leukemia factor* | 217 |
| 4. *5-methylcytosine G/T mismatch* | 190 |
| 5. *MINUTE protein* | 400 |
| 6. *Cyclin A* | 256 |
| 7. *Leucine-rich repeat protein* | 213 |
| 8. *2-cys –thioredoxin peroxidase* | 233 |
| 9. *Hydroxyacyl-CoA-dehydrogenase* | 244 |
| 10. *Carbamoyl phosphate synthetase* | 262 |
| 11. *Testes development-relate NYD-SP19* | 193 |
| 12. *Integrin beta 4 binding protein* | 166 |
| 13. *Cop 9 constitutive photomorphogenic homolog subunit 6* | 360 |
| 14. *Dihydropteridine reductase* | 327 |
| 15. *Casein kinase II beta chain* | 320 |
| 16. *Multicatalytic endopeptidase* | 217 |
| 17. *Innexin 2* | 208 |
| 18. *Tetrasparin D 107* | 316 |
| 19. *HLA-B associated transcript 1A, nuclear RNA helicase bata 1* | 150 |
| 20. *Heat shock protein 10 (HSP10)* | 176 |
| 21. *Muskelin 1, intracellular mediator containing kelch motifs* | 346 |
| 22. *Nonclatirin coat protein zeta 2* | 203 |
| 23. *Nit protein 2* | 292 |
| 24. *Endothelial cell growth factor 1* | 243 |
| 25. *Clatrin adaptor protein AP 50* | 267 |
| 26. *Calponin I* | 316 |

**Table 3.3** (cont.)

| Gene homologues | Expected size.(bp.) |
|---|---|
| 27. *Defender against cell death 1* | 214 |
| 28. *Receptor activating protein kinase C* | 187 |
| 29. *Vacuolar type H+ ATPase subunit A* | 150 |
| 30. *Death box protein 15* | - |
| 31. *ATP-dependent RNA helicase* | - |
| 32. *Splicing factor 3a,subunit 1* | 200 |
| 33. *DNA primase* | 256 |
| 34. *FK 506 binding protein 4* | 446 |
| 35. *ZZZ3* | 315 |
| 36. *Heterotrimeric GTP binding protein* (H) alpha subunit G-alpha-q | 216 |
| 37. *Keratinocyte associated protein 2* | 318 |
| 38. *PeF protein with along N-terminal hydrophobic domain* | 314 |

**Table 3.4** Gene homologues exhibiting non-specific amplification products from RT-PCR analysis

| Gene | Expected size (bp) |
|---|---|
| 1. *Glutathione peroxidase* | 145 |
| 2. *Serine proteinase inhibitor* | 293 |
| 3. *Leukemia virus receptor* | 230 |
| 4. *ETS1 protein* | 335 |
| 5. *COG4122: Predicted ;O-methyl transferase* | 399 |

**Table 3.5** Gene homologues that were not successfully amplified by RT-PCR

| Gene | Expected size (bp) |
|---|---|
| 1. *Phospholipase C* | 188 |
| 2. *Adenosylhomocysteinase* | 314 |
| 3. *Mitocondrial oxodicarboxylate* | 320 |
| 4. *Ras interacting protein RIPA* | 263 |
| 5. *Fructose 1,6-bisphosphate aldolase* | 170 |
| 6. *High mobility group protein DSP1* | 285 |
| 7. *Domain family member* | 318 |
| 8. *TRAP-like protien precursor* | 198 |
| 9. *C-myc bindig protein (AMY-1)* | 320 |
| 10. *Translocon associated protein gamma* | 266 |
| 11. *Cyclophilin 18* | 278 |
| 12 *mRNA splicing factor (deahbox)* | 278 |
| 13. *Cdc2 homologue* | 305 |
| 14. *Microspherule protein 1* | 317 |
| 15. *Cystathionine gamma lyase* | 145 |
| 16. *Splicing factor 34F* | 200 |
| 17. *NADPH-ferrihemoprotein reductase; NADPH-cytochroome P450 reductase* | 189 |
| 18. *Spermtail specific protein mst 101* | 301 |
| 19. *Nm2 protein* | 365 |
| 20. *Tetratricopeptide repeat domain 5* | 227 |
| 21. *Translationally controlled tumor protein* | 182 |
| 22. *Interleukin 1 receptor like 1 ligand precursor* | 162 |
| 23. *Ornithine decarboxylase* | 255 |
| 24. *Programmed cell death protein6* | 303 |
| 25. *Charperon subunit 8* | 323 |
| 26. *Presenilin enhancer* | 244 |
| 27. *Phosphoglucose isomerase* | 200 |

**Table 3.5** (cont.)

| Gene | Expected size (bp) |
|---|---|
| 28. *Dendritic cell protein* | 185 |
| 29. *Vitellogenin* | |
| 30. *Pre B cell colony enhancing factor* | 239 |
| 31. *Methyl CpG binding protein 2* | 206 |
| 32. *Synaptic vesicle-associated integral membrane protein* | 187 |
| 33. *Peroxinectin* | 366 |
| 34. *Cyclic AMP regulated protein* | 264 |
| 35. *Chloride intracellular channel 6-like protein* | 191 |
| 36. *Finger protein* | 378 |
| 37. *Exocyst complex component sec 6* | 305 |
| 38. *Small ubiquitin-like modifier* | |
| 39. *Prophenoloxidase activating factor (Propo factor)* | 154 |
| 40. *Heterogeneous nuclear ribonucleoprotein 87F* | 298 |
| 41. *Cyclophilin* A | 310 |
| 42. *Diphenoloxidase A2* | 205 |

**Table 3.6** Expression levels of gene homologues in various tissues of *P. monodon* broodstock

| Gene | Expression level | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | OV | TT | HE | HC | LO | IT | GL | PL | TG | ST | ES | HP |
| 1. *Female sterile* | +++ | | | | | | | | | | | |
| 2. *ATP/GTP binding protein* | +++ | | + | + | + | + | + | | + | + | | |
| 3. *Proactivator polypeptide precursor* | +++ | | +++ | + | ++ | + | ++ | + | ++ | + | + | ++ |
| 4. *TATA box binding protein associated factor9* | +++ | + | | + | | + | | | | | | |
| 5. *Broad complex Z4 isoform* | +++ | | | ++ | + | + | + | + | + | + | + | + |
| 6. *Zonadhesin isoform 4* | +++ | + | + | + | + | | | | | | | |
| 7. *Ovarian lipoprotein receptor* | +++ | | | | | | | | | | | |
| 8. *Small androgen receptor interacting protein* | +++ | | + | + | + | | | | | | + | |
| 9. *NADP dependent leukotriene B4 12 hydroxy dehydrogenase* | +++ | + | ++ | + | + | + | | + | + | | | |
| 10. *Aminopeptidase* | +++ | | + | + | + | + | + | | | + | | |
| 11. *Y-box protein* | +++ | ++ | + | ++ | | + | ++ | ++ | + | + | + | + |
| 12. *Wolf hirschorn syndrome candidate 1 protein* | +++ | | + | | + | + | | | | | | |
| 13. *RAB protein 10 CG 17060-PA* | +++ | + | ++ | ++ | ++ | + | ++ | + | + | + | + | |

**Table. 3.6** (cont.)

| Gene | Expression | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | OV | TT | HE | HC | LO | IT | GL | PL | TG | ST | ES | HP |
| 14.*Chromobox protein* | +++ | + | + | ++ | + | + | + | + | + | + | | |
| 15 *Aspartate aminotransferase* | +++ | + | + | ++ | | + | | ++ | | | + | + |
| 16.*Endothelial cell growth factor 1* | + | + | ++ | + | + | + | + | | | | | |
| 17. *COG4122: Predicted ;O-methyl transferase* | ++ | + | + | +++ | + | + | + | | + | + | | ++ |
| 18.*3-Oxoacid CoA transferase* | ++ | + | + | | | | | + | | ++ | | |
| 19.*Cyclin A* | +++ | + | | + | + | | + | + | | + | + | |
| 20. *Dolichyl diphosphooligosaccharide protein glycotransferase* | +++ | ++ | + | ++ | + | + | + | + | | + | + | + |
| 21.*Nuclear autoantigenic sperm protein* | +++ | + | | | | + | | | | | | |
| 22. *Asparaginyl tRNA synthetase* | +++ | ++ | | ++ | + | + | + | + | | +++ | + | |
| 23 *CG-1681-PA* | +++ | + | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ | + | + |
| 24. *RuvB like protein2* | +++ | + | ++ | ++ | ++ | ++ | ++ | ++ | ++ | ++ | + | ++ |
| 25.*Phosphatidyl serine receptor short form* | +++ | + | ++ | ++ | ++ | + | + | + | + | + | + | |
| 26.*Adipose differentiation related protein* | +++ | | + | + | + | | | | | + | | |

+++ = high expression level, ++ = moderate expression level, + = low expression level, - = not expressed in a particular tissue.

### 3.2 Identification of polymorphic SSCP patterns of cDNA genes of *P. monodon* exhibiting different GSI

RT-PCR of 22 gene homologues including those (5) expressed only in ovaries but not testes of *P. monodon* broodstock (*adipose differentiation related protein, ovarian lipoprotein receptor, female sterile, broad complex Z4 isoform* and *proactivator polypeptide precursor*) and those (17) exhibiting preferential expression in ovaries than testes (*TATA binding protein associated factor9, RAB protein 10 CG 17060-PA, nuclear autoantigenic sperm protein, dynein heavy chain, ESO 3 protein, dolichyl diphosphooligocharide protein glycotransferaase, survival motor neuron, 3-oxoacid CoA transferrase, thiolase, ferritin, Y-box protein p-50, tissue specific transplantation antigen p35B like protein, asparaginyl tRNA synthetase, calcineurinB, cop 9 subunit 6, immonophilin FKBP 52* and *tetrasparin D 107*) were carried out using the first strand cDNA of ovaries or testes of *P. monodon*. The RT-PCR product was further analyzed by SSCP.

*Ovarian lipoprotein receptor* which is specifically expressed in ovaries but not other tissues displayed highly polymorphism among individuals of *P. monodon* (Figure 3.22). Nonsynonymous mutations of this gene should be further identify in different shrimp as they may relate with functional activity of *ovarian lipoprotein receptor* during veltellogenesis.

On the other hand, the RT-PCR product of *female sterile* (Figure 3.23) and *adipose differentiation related protein* (Figure 3.24) was the not polymorphic across shrimp exhibiting different stages of ovarian development.

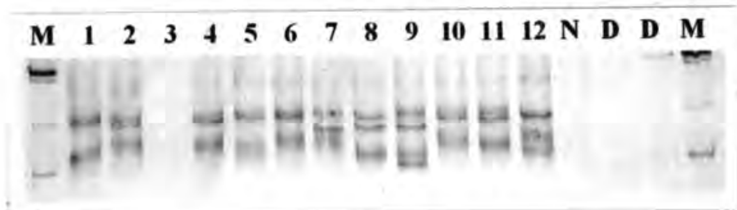



**Figure 3.22** The amplified cDNA products of *ovarian lipoprotein receptor* from ovaries of *P. monodon* having the GSI of 5.69, 4.70, 3.02, 2.40, 2.13, 2.02, 1.90, 1.43, 1.10, 0.92, 0.87, and 0.65 corresponding to lanes 1 - 12 were analyzed by SSCP. Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

Although polymorphic the RT-PCR products of *immonophilin FKBP52* was observed, identified isotypes of *immonophilin* were found in both ovaries and testes of *P. monodon* (Figure 3.25).

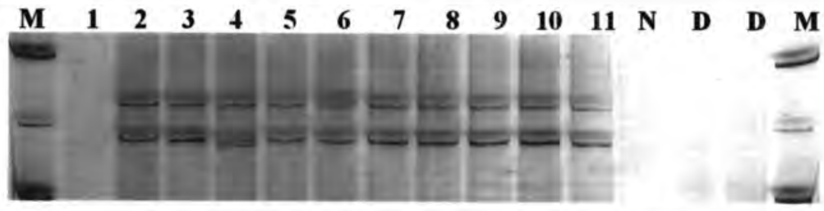



**Figure 3.23** The amplified cDNA products of *female sterile* of *P. monodon* having the GSI of 5.69, 4.70, 3.02, 2.13, 2.02, 1.90, 1.43, 1.10, 0.92, 0.87, and 0.65 corresponding to lanes 1 - 11 were analyzed by SSCP. Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

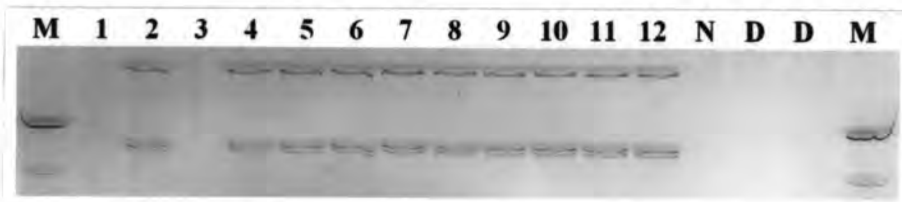



**Figure 3.24** The amplified RT-PCR products of *adipose differentiation related protein* from ovaries of *P. monodon* having the GSI of 5.69, 4.70, 3.02, 2.13, 2.02, 1.90, 1.43, 1.10, 0.92, 0.87, and 0.65 corresponding to lanes 1 - 12 were analyzed by SSCP. Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

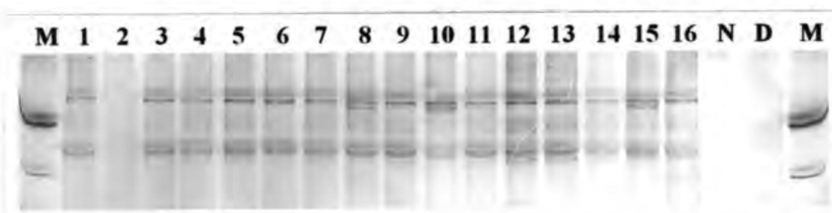



**Figure 3.25** The amplified RT-PCR products of *immunophilin FKBP52* from testes of male *P. monodon* having the GSI of 1.11, 0.84, 0.76, 0.67, 0.61, 0.58, 0.54, and 0.52and ovaries of female *P. monodon* having the GSI of 3.02, 2.02, 1.90, 1.43, 1.10, 0.92, 0.87, and 0.65 corresponding to lanes 1 – 8 and 9 – 16, respectively were analyzed by SSCP.Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

*Asparaginyl t-RNA synthetase* is preferentially expressed in ovaries of *P. monodon*. Different SSCP isotypes were found in testes but did not correlate with the GSI values. In contrast, additional SSCP isotypes of *asparaginyl tRNA synthetase* were found in the female shrimp possessing the GSI of 1.90, 2.02 and 2.13 compared with those having the GSI between 0.65 – 1.43% (Figure 3.26). This indicated that rapid accumulation of proteins containing Asp is required during vitellogenesis.

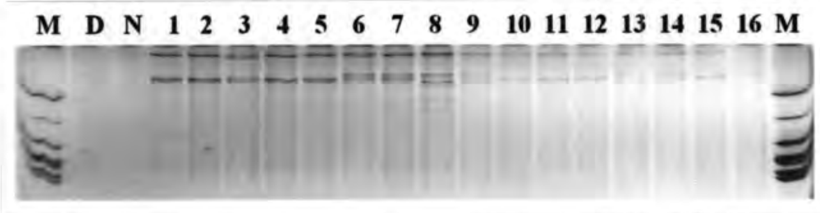



**Figure 3.26** The amplified RT-PCR products of *asparaginyl tRNA synthetase* from ovaries of female *P. monodon* having the GSI of 0.65, 0.87, 0.92, 1.10, 1.43, 1.90, 2.02, and 3.02 and testes of male *P. monodon* having the GSI of 0.52, 0.54, 0.58, 0.61, 0.67, 0.76, 0.84, and 1.11 corresponding to lanes 1 – 8 and 9 – 16, respectively. were analyzed by SSCP. Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

A fixed SSCP pattern of *TATA binding protein associated factor9* was found in ovaries and testes of *P. monodon* (Figure 3.27). This suggested sex-specific isotypes of *TATA binding protein associated factor9*. It is interesting to examine SSCP of the genomic DNA of *TATA binding protein associated factor9* for the possibility to develop genomic sex determination of *P. monodon*.

The RT-PCR product of *nuclear autoantigenic sperm protein* was highly polymorphic whereas that of testes was less polymorphic that of ovaries. Individual-specific SSCP patterns were observed in ovaries of *P. monodon*. Correlations between single nucleotide polymorphism (SNP) of this gene and the quality of oocytes should be further examined.

Limited polymorphism was found from RT-PCR products of *ferritin* (Figure 3.29) and *3-oxoacid CoA transferase* (Figure 3.30) and *ESO-3 protein* (Figure 3.31). *Tissue-specific transplantation antigen p35B like protein* (Figure 3.32) revealed slightly greater polymorphism than *ferritin* and *3-oxoacid CoA transferase*. Nevertheless, *tetrasparinD 107* was highly polymorphic in both ovaries and testes (Figure 3.33).

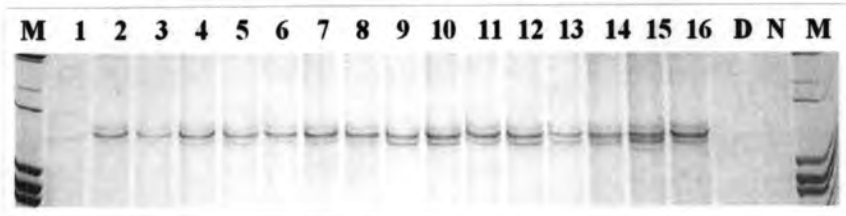



**Figure 3.27** The amplified cDNA products of *TATA binding protein associated factor9* from testes of male *P. monodon* having the GSI of 1.11, 0.84, 0.76, 0.67, 0.61, 0.58 , 0.54, and 0.52 and ovaries of female *P. monodon* having the GSI of 3.02, 2.13, 2.02, 1.90, 1.43, 1.10, 0.92 ,0.87, and 0.65 corresponding to lanes 1 – 8 and 9 – 16, respectively were analyzed by SSCP. Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

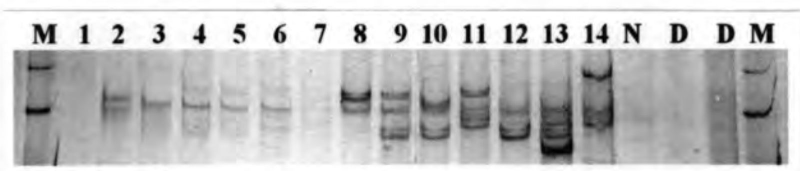



**Figure 3.28** The amplified RT-PCR products of *nuclear autoantigenic sperm protein* from testes of male *P. monodon* having the GSI of 1.11, 0.84, 0.76, 0.67, 0.61, 0.58, , and 0.54 and ovaries of female *P. monodon* having the GSI of 4.70, 2.13, 2.02 ,1.90, 1.43, 0.92, and 0.87 corresponding to lanes 1 - 7 and 8 - 14, respectively we re analyzed by SSCP. Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

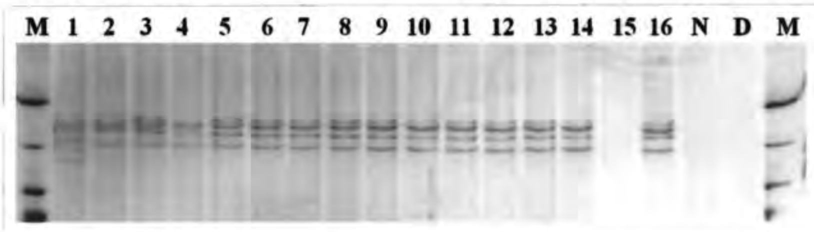



**Figure 3.29** The amplified cDNA products of *ferritin* from testes of male *P. monodon* having the GSI of 1.11, 0.84, 0.76, 0.67, 0.61, 0.58, 0.54, and 0.52 and ovaries of female *P. monodon* having the GSI of 4.70, 3.02, 2.02, 1.90, 1.43, 0.92, 0.87, and 0.65corresponding to lanes 1 – 8 and 9 – 16, respectively were analyzed by SSCP. Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

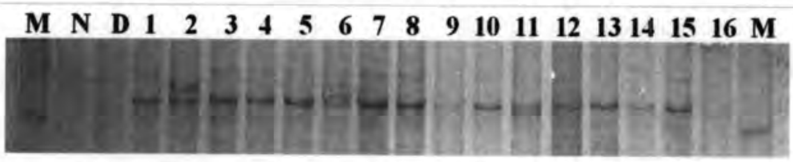



**Figure 3.30** The amplified cDNA products of *3-oxoacid CoA transferase* from ovaries of female *P. monodon* having the GSI of 0.65, 0.87, 0.92, 1.43, 1.90, 2.02, 2.13 and 4.70 and testes of male *P. monodon* having the GSI of 0.52, 0.54, 0.58, 0.61, 0.67, 0.76, 0.84, and 1.11 and corresponding to lanes 1 – 8 and 9 – 16, respectively were analyzed by SSCP Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.


**Figure 3.31** The amplified cDNA products of *ESO-3 protein* from ovaries of female *P. monodon* having the GSI of 0.65, 0.87, 0.92, 1.43, 1.90, 2.02, 2.13 and 4.70 and testes of male *P. monodon* having the GSI of 0.52, 0.54, 0.58, 0.61, 0.67, 0.76, 0.84, and 1.11 and corresponding to lanes 1 – 8 and 9 – 16, respectively were analyzed by SSCP Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

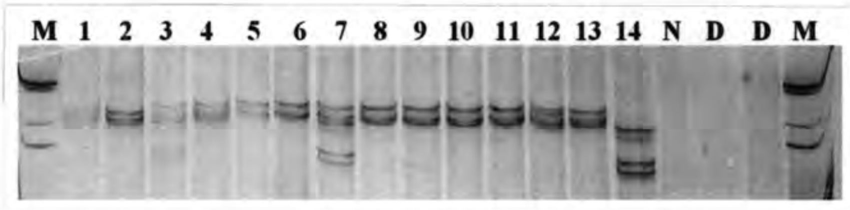



**Figure 3.32** The amplified cDNA products of *tissue-specific transplantation antigen p35B like protein* from testes of male *P. monodon* having the GSI of 1.11, 0.84, 0.76, 0.67, 0.61, 0.58, and 0.54 and ovaries of female *P. monodon* having the GSI of 4.70, 2.13, 2.02, 1.90, 1.43, 0.92, and 0.87 corresponding to lanes 1 – 7 and 8 – 14, respectively were analyzed by SSCP Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

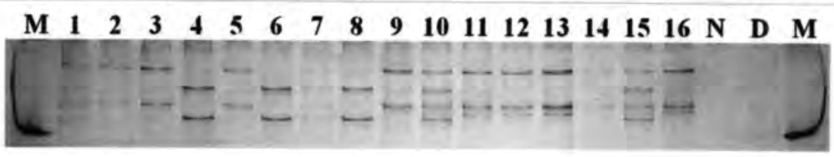



**Figure 3.33** The amplified cDNA products of *tetrasparinD 1073* from testes of male *P. monodon* having the GSI of 1.11, 0.84, 0.76, 0.67, 0.61, 0.58, 0.54, and 0.52 and ovaries of female *P. monodon* having the GSI of 4.70, 2.13, 2.02, 1.90, 1.43, 0.92, 0.87 and 0.65 corresponding to lanes 1 – 8 and 9 – 16, respectively were analyzed by SSCP Lanes N, M and D are the negative control, a 100 bp DNA marker and the non-denatured PCR product (double strand control), respectively.

### 3.3 Isolation and characterization of full length cDNA using RACE-PCR

#### 3.3.1 RNA extraction and first strand synthesis

The quantity and quality of total RNA was spectrophotometrically and electrophoretically examined, respectively. The ratio of $OD_{260}/OD_{280}$ of extracted RNA ranged between 1.8 – 2.0 indicating that the extracted RNA was relatively pure. Agarose gel electrophoresis indicated smear total RNA with a few discrete bands implying the accepted quality of extracted total RNA (Figure 3.34). The ovarian mRNA was purified and large amount of mRNA was obtained. The purified mRNA was subjected to the synthesis of the 5′ and 3′RACE template

**Table 3.7** SSCP pattern of the RT-PCR product of various gene homologues

| Gene | Expected size (bp) | SSCP |
|---|---|---|
| 1. *Adipose differentiation related protein* | 267 | Monomorphism |
| 2. *Rab protein 10 CG 17060-PA* | 177 | Monomorphism |
| 3.*Nuclear autoantigenic sperm protein* | 301 | Polymorphism |
| 4.*Ovarian lipoprotein receptor* | 354 | Polymorphism |
| 5. *Dynein heavy chain* | 192 | Monomorphism |
| 6. *ESO 3 protein* | 349 | Polymorphism |
| 7. *Dolichyl diphosphooligocharide protein glycotransferaase* | 233 | Polymorphism |
| 8. *Survival motor neuron* | 245 | Monomorphism |
| 9. *3-oxoacid CoA transferrase* | 303 | Polymorphism |
| 10. *Thiolase* | 231 | Monomorphism |
| 11. *TATA binding protein associated factor9* | 320 | Polymorphism |
| 12. *Ferritin* | 280 | Polymorphism |
| 13. *Y-box protein* | 435 | Polymorphism |
| 14. *Tissue specific transplantation antigen p35B like protein* | 352 | Polymorphism |
| 15. *Asparaginyl tRNA synthetase* | 330 | Polymorphism |
| 16 .*Immonophilin FKBP 52* | 312 | Polymorphism |
| 17. *CalcineurinB* | 319 | Monomorphism |
| 18. *Female sterile* | 296 | Monomorphism |
| 19. *Proactivator polypeptide precursor* | 266 | Monomorphism |
| 20. *Cop 9 subunit 6* | 360 | Polymorphism |
| 21. *Broad complex Z4 isoform* | 192 | Monomorphism |
| 22. *Tetrasparin D 107* | 316 | Polymorphism |

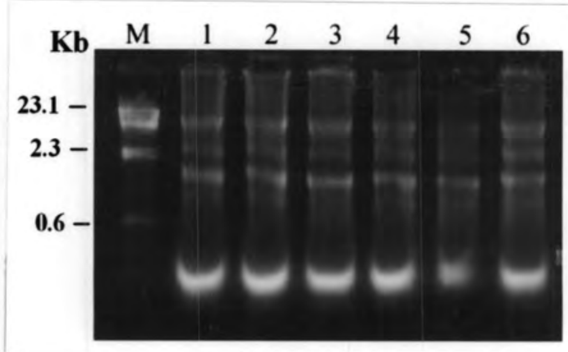



**Figure 3.34** A 0.8% ethidium bromide-stained agarose gel showing the quality of RNA from ovaries of different individuals of *P. monodon* (lanes = 1 - 6). Lanes M is λ/*Hin*d III marker.

### 3.3.2 Isolation and characterization of the full length cDNA of gene homologues

#### 1. *3-oxoacid CoA transferase*

A 1200 bp fragment was obtained from 5'RACE-PCR of a *3-oxoacid CoA transferase* primer (Figure 3.35). An amplification fragment was cloned and sequenced for both directions. Nucleotide sequences of EST and RACE-PCR were assembled and analyzed.

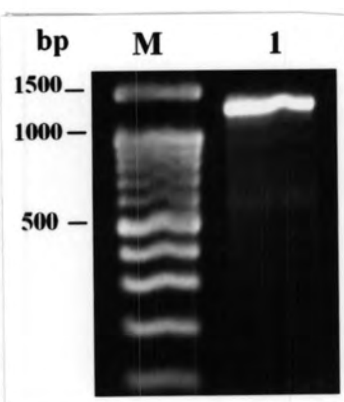
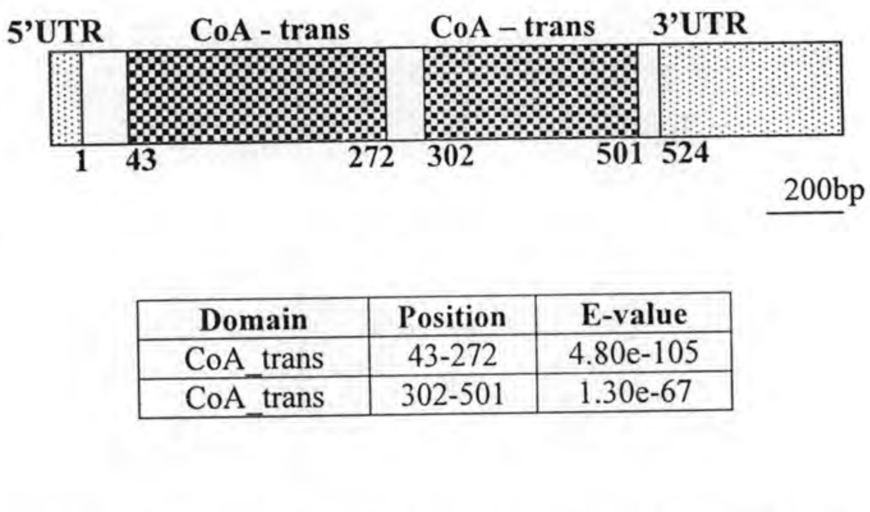



**Figure 3.35** The primary 5'RACE-PCR product of *3-oxoacid CoA transferrase* (lane 1). A 100 bp (lane M) DNA ladder was used as the markers.

The full length cDNA of *3-oxoacid CoA transferrase* of *P. monodon* was 2129 bp in length. The ORF of *3-oxoacid CoA transferrase* was 1575 bp encoding a polypeptide of 524 amino acids. The 5′ and 3′ UTRs of *3-oxoacid CoA transferrase* were 58 and 496 bp (excluding the poly A tail). The poly A additional signal (AATAAA) was located between 2087 - 2092 of the entire sequence. The closest species according to the best hit approach of the full length cDNA of *3-oxoacid CoA transferrase* of *P.monodon* was *Mus musculus* (E-value = 0.00).

The calculated pI and MW of *3-oxoacid CoA transferrase* of *P. monodon* was 5.57 and 56079.41 dalton, respectively. The signal peptide was not found in this putative nonsecretory protein. Two domains of the CoA transferrase domain functionally catalyzing the reversible transfer of CoA from one carboxylic acid to another, were found at the amino acid positions 43 - 272 (4.80e-105) and 302 - 501 (1.30e-67; Figure 3.36).




| Domain | Position | E-value |
|---|---|---|
| CoA_trans | 43-272 | 4.80e-105 |
| CoA_trans | 302-501 | 1.30e-67 |

**Figure 3.36** Diagram illustrating the full length cDNA of *3-oxoacid CoA transferrase* of *P.monodon*.The CoA transferrase domains were found in this transcript. The scale bar is 200 bp in length.

```
CTAATACGACTCACTATAGGGCAAGCAGTGGTTCAACGCAGAGTACGCGGGGAGCAACAT   60
                                                           M    1
GGCGTCGCGGGCGAGCGGAGGCGTCGGCCGATTCGCGAGGGTTTTGGTGTCCTCGCGCTC  120
 A  S  R  A  S  G  G  V  G  R  F  A  R  V  L  V  S  S  R  S    21
GCACAGGCTCTGCAAGGTTCAAGTTTGCTTGCTGTCAACAAGTATTCGCCGTGATGCCAA  180
 H  R  L  C  K  V  Q  V  C  L  L  S  T  S  I  R  R  D  A  K    41
GTTTTATGACAGCGCGCTGGAGGCTGTCGATGATATTCCCTCGGGGTCAAAGCTCCTTGT  240
```

F Y D S A L E A V D D I P S G S K L L V 61
CGGAGGCTTCGGTCTCTGCGGCATTCCGGAGAACCTGATTGGAGGGCTGCTGGAGACCAA 300
G G F G L C G I P E N L I G G L L E T K 81
AGTCAAGGACTTGACAGTGGTGAGCAACAACGCAGGAGTGGACAACTTTGGCCTGGGGCT 360
V K D L T V V S N N A G V D N F G L G L 101
CCTGCTGGCACAGAAGCAGATCAAGCGCATGATCTCCTCCTACGTGGGCGAAAATGCCGA 420
L L A Q K Q I K R M I S S Y V G E N A E 121
GTTCGAGCGGCAGTACCTGAGCGGGGAACTGGAGGTGGAGCTTACGCCGCAGGGCACGTT 480
F E R Q Y L S G E L E V E L T P Q G T L 141
GGCCGAGCGATGTCGTGCAGGGGGGGCTGGCATCCCTGCCTTCTTCACTCCCACGGGTTT 540
A E R C R A G G A G I P A F F T P T G F 161
TGGCACACTCGTCCACGAGGGAGGGTCTCCCATCAAATATGGGGAAGGTGGTGCTATTCA 600
G T L V H E G G S P I K Y G E G G A I Q 181
GATCCAGAGTGCCCCAAGGGAGAGCCGGATATTCAATGGACGGAACTACATCATGGAGGA 660
I Q S A P R E S R I F N G R N Y I M E E 201
GGCCATTACTGGGGACTTTGCACTCATCAAGGCGTGGAAGGCCGACCGGGCGGGCAACCT 720
A I T G D F A L I K A W K A D R A G N L 221
TCTCTTTAGGAAGACAACACGCAACTTCAACCTGCCGATGTGCAAGGCTGCCAAGACCAC 780
L F R K T T R N F N L P M C K A A K T T 241
CATAGTTGAAGTGGAGGAGGTCGTCGACATTGGGGAAATCCCAGAGGACAGCGTACATGT 840
I V E V E E V V D I G E I P E D S V H V 261
CCCTGCCATCTATGTGGACCGCATCATTACAGGGGAGAAGTATGAGAAGAGAATTGAGCG 900
P A I Y V D R I I T G E K Y E K R I E R 281
CCTGACCCTGCGCAAGGAGAAGAAGAAGAGTGCCGCATCCTCCAGCCCGGCCGTGGCCAT 960
L T L R K E K K K S A A S S S P A V A M 301
GAGGGAGCGCATTGTGCGACGCGCAGCCCTCGAGTTCAAAGACGGCATGTACGCCAACCT 1020
R E R I V R R A A L E F K D G M Y A N L 321
GGGGATTGGGATGCCCATGCTTGCCAGCAATTACATCCCTGATGGCACGAATGTGCAGCT 1080
G I G M P M L A S N Y I P D G T N V Q L 341
GCAGAGTGA***GAACGGGGTGCTGGGCTTGGGT***CCCTTCCCTGCCCCAGGTGAGCAAGATCC 1140
Q S E N G V L G L G P F P A P G E Q D P 361
TGACCTCATCAACGCTGGCAAAGAGACCGTGACTGTCACGCCTGGGGCCTCCTACTTTGG 1200
D L I N A G K E T V T V T P G A S Y F G 381
CTCTGACGAGAGCTTTGCCATGATCCGAGGTGGCCACGTGGACCTGACAATCCTTGGGGC 1260
S D E S F A M I R G G H V D L T I L G A 401
CATGGAAGTCTCTCAGTACGGTGACCTTGCCAACTGGATGATTCCGGGCAAGATGGTGAA 1320
M E V S Q Y G D L A N W M I P G K M V K 421
AGGCATGGGTGGTGCAATGGACCTGGTGTCATCACCTGGCACGAAGGTGGTGGTGACAAT 1380
G M G G A M D L V S S P G T K V V V T M 441
GGAGCACTCGGCCAAGAAGGGTGGACATAAGATCGTGGAAGCCTGCTCGCTCCCCCTCAC 1440
E H S A K K G G H K I V E A C S L P L T 461
AGGCAAGAACTGTGTCGATATGATAATCACAGAGAAGTGCGTATTCAGTGTAGACAAGGA 1500
G K N C V D M I I T E K C V F S V D K E 481
GAATGGCCTGACTCTGGTTGAAATCGCCGATGGCGTTACCATCGAGGACGTCGTATCCAG 1560
N G L T L V E I A D G V T I E D V V S S 501
CACAGGATGCTTATTTGAAGTGGCGGAAAACCTCAAGCCAATGGGACAGATTGAGGTTGC 1620
T G C L F E V A E N L K P M G Q I E V A 521
AGATGAA**TAA**AGGCTGAAATTTGCTTAGACTTAGTCATCAAGTCAAAAGAAATATAGAT 1680
D E * 523
ATTCAGATTAGTAGAGTGTGTCATTGTGCACATGGTTTGCTGGCACTTTTGATTTAACTT 1740
TTGATCTGAACCTTTGATTAACTAATGGGCATATTTGAGCCCGAGGCATTAATTATAAA 1800
GGGATTCACGTATCCTTTTGCTACTATAGTTCTTCCACACCCATACTAACAGCACGATAA 1860
TTTTGCATAAAGAAAGAAGTGCAGATAATAAATTATGATAATAAATAAAGTTTTTACTCT 1920
ATTCAAGCCAAGGATTCCTTCTTTCTTGGATCATTGCCTTAGTCATTGTAGTGGCATTCA 1980
GTTTTACTCTTGTGTGACAGTGTGTTAAAAGCAAGTTTTTCTTGAGCTTACATCATGTGT 2040
TCAGCTTCAGTTGATTGTCTGCACAACATACAAAAGAAAAGGAAAA**AATAAA**ACAAAAAG 2100
CTTTATCATGAAAAAAAAAAAAAAAAAAA 2129

**Figure 3.37** The full length cDNA sequences of *3-oxoacid CoA transferrase* of *P. monodon*. Start and stop codons are illustrated in boldface and underlined. The 5' RACE-PCR primer is underlined, boldfaced and italicized. The polyA additional signal is boldfaced.

### 2. *Adipose differentiation related protein* (*PMADRP*)

A 600 bp fragment was obtained from 5′RACE-PCR of a *PMADRP* primer (Figure 3.38). An amplification fragment was cloned and sequenced for both directions. Nucleotide sequences of the original EST and 5′RACE-PCR were assembled and analyzed.

The full length cDNA of *PMADRP* was 1336 bp in length where the 5′ and 3′UTRs of *PMADRP* were 78 and 148 bp (excluding the poly A tail), respectively. The ORF of *PMADRP* was 1110 bp encoding a polypeptide of 369 amino acids. The poly A additional signal (AATAAA) was located between 1302 - 1307 of the entire sequence of *PMADRP* (Figure 3.40).

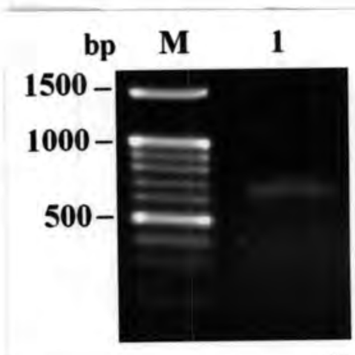
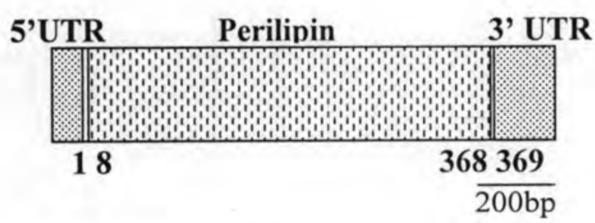



**Figure 3.38** The primary 5′RACE-PCR product of *PMADRP* (lanes A). A 100 bp DNA ladder (lanes M) was used as the markers.




| Domain | Position | E-value |
|---|---|---|
| Perilipin | 8 – 368 | 3.96e-06 |

**Figure 3.39** Diagram illustrating the full length cDNA of *PMADRP*. The perilipin domain was found in this transcript.

The closest sequence to *PMADRP* was *adipophilin* (an alternative name of *adipose differentiation-related protein , ADRP*) of *Canis familiaris* (E-value = 9e-18).

The calculated pI was 6.90 with the molecular weight of 39729.22 dalton. The signal peptide was not found in this putative non-scretory protein. A *perilipin* domain was found in the ORF at positions 8 - 368 (3.90e-06; Figure 3.39). Perilipin is a modulator of adipocyte lipid metabolism and adipophilinis which involve in the development and maintenance of adipose tissues.

```
CTAATACGGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGTCTCT     60
GCCCAGTTTTCGAACGAAATGGCTCCGAACACGAGTAATGCGACACCGGCTGCAGAAGGC    120
                  M  A  P  N  T  S  N  A  T  P  A  A  E  G      14
TTTTTTGAGAGAGTCTTGCTCTTGCCTGTGCTTAGCGACGCCATCACCATCGTTTCTCAT    180
 F  F  E  R  V  L  L  L  P  V  L  S  D  A  I  T  I  V  S  H     34
GCCTACAAAATTACCCAAGACCGATACCAGTATGTGGGAACTGCTTTGCGAGTAGCAGAG    240
 A  Y  K  I  T  Q  D  R  Y  Q  Y  V  G  T  A  L  R  V  A  E     54
GCAGGCATCCGGGTGGCCACCGAGGGAGCTCTCCCTCTGGCTATGCCCCTTCTGCACCCC    300
 A  G  I  R  V  A  T  E  G  A  L  P  L  A  M  P  L  L  H  P     74
CTCGTGGACCGCGTCGGAGGATGGAGTACCCTGGATGAGTGGGCATGTCGCGGTCTGGAT    360
 L  V  D  R  V  G  G  W  S  T  L  D  E  W  A  C  R  G  L  D     94
CGCGTTGAGGAGGCAGCGCCCATCATAACCAAGTCAACGGATGAGATCGTGAGTTCAGCG    420
 R  V  E  E  A  A  P  I  I  T  K  S  T  D  E  I  V  S  S  A    114
CGCCGTCGGGTGCTCAGCGTCGTGGCGGGGAAGGACGCTCTACCCCCCTCCCTCTCCGCC    480
 R  R  R  V  L  S  V  V  A  G  K  D  A  L  P  P  S  L  S  A    134
GCCGTCACGTCGCGGGCCAACGACACCGTTGACGTCATCGCTGCGAGTCGGGGCGGGCGA    540
 A  V  T  S  R  A  N  D  T  V  D  V  I  A  A  S  R  G  G  R    154
GTCGTCGCGGGCGCTGCGGAGCGGGTGCTCAACACTGCCCACACGCTGGTCGACGCCTAC    600
 V  V  A  G  A  A  E  R  V  L  N  T  A  H  T  L  V  D  A  Y    174
TTACCACCCCGCGAGGGCGACCTTCACGACACAGATGGCGCGACGCAACGGTGGGCGTA    660
 L  P  P  R  E  G  D  L  H  D  T  D  G  R  D  A  T  V  G  V    194
AAGGTGATAGCGCTGGCAAGCAAGACGAGGCGGCGTTTGGCCCGCGCGCTTCATTCGCTC    720
 K  V  I  A  L  A  S  K  T  R  R  R  L  A  R  A  L  H  S  L    214
GCGCATCCGCACCACGCCCACGCCCACACCAACGCCGACGCCTGTGACGCCACGGATGCA    780
 A  H  P  H  H  A  H  A  H  T  N  A  D  A  C  D  A  T  D  A    234
TCGGGCAGCTTCCTGATTGCCTTCAGTCGGGAGTGTGTGTCTGGCTGGCACTCTGAAGTC    840
 S  G  S  F  L  I  A  F  S  R  E  C  V  S  G  W  H  S  E  V    254
ACGCGGCAGCCGGAGCCCCACGAGGCGGTGCCTGTAGTTTTGAGAATCGCACGAACTTCC    900
 T  R  Q  P  E  P  H  E  A  V  P  V  V  L  R  I  A  R  T  S    274
TACAGGTACCTAAGGAACGTTACGGAGAATCTTGGATCACTCGTTGTTGCTGTTCGCAAT    960
 Y  R  Y  L  R  N  V  T  E  N  L  G  S  L  V  V  A  V  R  N    294
TCCCTGGCTCGGAACGCAGTGAGAGAAGCCTTGAGCGGCGCGGCGGCCAACGGATGTTA    1020
 S  L  A  R  N  A  V  R  E  A  L  S  G  A  A  A  N  G  C  Y    314
GTCGTTTCCGAGATGCAAGCCTGGGGAGAGTACTTTGTTGTTGTCATGTCGATGACGCCA   1080
 V  V  S  E  M  Q  A  W  G  E  Y  F  V  V  V  M  S  M  T  P    334
AGTCTGATCGAAGAGGCTTCGCTTCGCTCCAGGGTGTGGGCTGAAGGGTGTCTGGAATCT   1140
 S  L  I  E  E  A  S  L  R  S  R  V  W  A  E  G  C  L  E  S    354
GTAATGAAAATGAAGTATACAAATCCACATCTGTGTTGGCAGTGACTGACTGACGCCAT    1200
 V  M  K  M  K  Y  T  N  P  H  L  C  W  Q  *                   368
CAGTGAAGACCAGAGGAACTCCCTGAATTGTTTTTCACGGCTGGATAATTACTCTTAATC   1260
GATGTTAGCTGTTTTGTGGTTTGGACGTTTGTAATAAGGGTAATAAAATTACTGAAAGCA   1320
AAAAAAAAAAAAAAAAA                                              1337
```

**Figure 3.40** The full length cDNA sequences of *ADRP* of *P. monodon*. Start and stop codons are illustrated in boldface and underlined. The 5'RACE-PCR primer is underlined, boldfaced and italicized. The polyA additional signal is boldfaced.

### 3. *Asparaginyl tRNA synthetase (PMATRS)*

A 980 bp fragment was obtained from 5'RACE-PCR of a *PMATRS* primer (Figure 3.41). An amplification fragment was cloned and sequenced for both directions. Nucleotide sequences of the original EST and 5'RACE-PCR (both direction + 1 internal primer for the primer walking) were assembled and analyzed.

The full length cDNA of *PMATRS* was 1835 bp in length where the 5' and 3'UTRs of *PMATRS* were 106 and 63 bp (excluding the poly A tail), respectively. The ORF of *PMATRS* was 1686 bp encoding a polypeptide of 561 amino acids. The poly A additional signal (AATAAA) was located between 1803 - 1808 of the entire sequence of *PMATRS* (Figures 3.42 and 3.43).

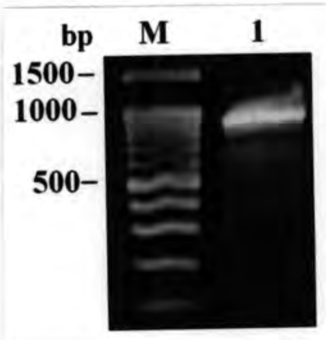
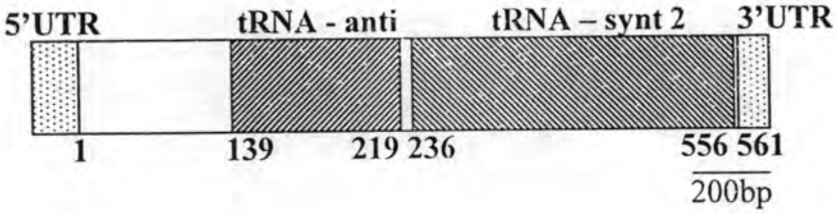



**Figure 3.41** The primary 5'RACE-PCR product of *PMATRS* (lane 1). A 100 bp (lanes M) DNA ladder was used as the markers

The closest sequence to *PMATRS* was *asparaginyl-tRNA synthetase* of *Aedes aegypti* (E-value = 0.0). The calculated pI and MW of pI *PMASP* were 6.00 and 64058.09 dalton, respectively with no signal peptide in this putative non-secretory protein. An anti -codon binding domain (positions 139 - 219; 9.10e-15) and tRNA synthetases class II domain (positions 236 - 556; 3.70e-76) were found (Figure 3.43).

The anti-codon binding domain I (containing the OB domain) binds to nucleic acids and catalyzes the addition of an amino acid to the appropriate tRNA molecule. The tRNA synthetases class II domain catalyze the attachment of an amino acid to its cognate transfer RNA molecule in a highly specific two-step reaction.




| Domain | Position | E-value |
|---|---|---|
| tRNA_anti | 139-219 | 9.10e-159 |
| tRNA-synt_2 | 236-556 | 3.70e-76 |

**Figure 3.42** Diagram illustrating the full length cDNA of *PMATRS*. The anti-codon binding domain and tRNA synthetase class II domain were found in this transcript. The scale bar is 200 bp in length

TTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCACAGTACGCGGGGAGTGTTA 60
CACAGAACCTCATCGAGTGCTGACGTCGCGCTTCCTTTCTCCCGGA**ATG**GCAAACGAGAC 120
M A N E T 5
GACAGAACAAATGGCTCAAATGAGTTTAGGGGAAATTTACACCTCGGAGAAACATGGCTC 180
T E Q M A Q M S L G E I Y T S E K H G S 25
AGATGAGACGGGGGACGGCACTGCTGACCGACCCTTGAAGACGGCCATGCAGGCCATGCG 240
D E T G D G T A D R P L K T A M Q A M R 45
TAAGGCTGGCAAGGAGTCTTTCCCCACCATCTACGTTGACGCCAAAGAGGAAGGACAGAA 300
K A G K E S F P T I Y V D A K E E G Q K 65
ATATGAAGTGATTGCAAAGAGTCAGTTGAAGAAGCTCACAAAACTATGGAAGGATGAGTG 360
Y E V I A K S Q L K K L T K L W K D E C 85
CAAAAAGATCGAAGCTCGGCTCAAGAAAGAGAAGGAGGATGCCGAGAAGAGAGCAAAGGC 420
K K I E A R L K K E K E D A E K R A K A 105
CATAGAGGAAGCCAAGAAGGTCGTGATCACAGAGGACAAGAGCCTGCCAGCCCCCGTCTG 480
I E E A K K V V I T E D K S L P A P V C 125
CATCAAGATCCGCGACGGAAAGGAGCACAGAGGGAAACGTGTCAAAATCAGGGGCTGGGT 540
I K I R D G K E H R G K R V K I R G W V 145
CCACCGTCTGAGGAGGCAGGGCAAAAACATGATGTTCATCGTTCTGCGAGATGGCTCAGG 600
H R L R R Q G K N M M F I V L R D G S G 165
ATACCTACAGTCGGTGCTGACGGACCAGCTTTGCCAGACGTACGAGGCCATCATCCTCAA 660
Y L Q S V L T D Q L C Q T Y E A I I L N 185
CACTGAAAGCACCGTCACTCTCTATGGCATGCTCCAGGAGGTGCCGGAGGGGAAGGAGGC 720
T E S T V T L Y G M L Q E V P E G K E A 205
CCCAGGTGGACACGAACTCCAGGTAGATTACTGGGAGCTGGTGGGAGAGTCACCAGCTGG 780
P G G H E L Q V D Y W E L V G E S P A G 225
TGGTGCCGAGGCAGAGATCAACCAGCTGTCCAACCCTGATGTGCAGCTGGACAAGCGACA 840
G A E A E I N Q L S N P D V Q L D K R H 245
CCTCATGATCCGTGGAGAGAACTG***CTCTAAGGTGCTGCGGCTGCGTT***CAATCCTTATGAA 900
L M I R G E N C S K V L R L R S I L M K 265
GGCCTTCGTGGACCACTACACGGACCGCGGCTACGAGTGGATATCCCCCCCGACCCTGGT 960
A F V D H Y T D R G Y E W I S P P T L V 285
GCAGACGCAGTGTGAGGGAGGCTCCACGCTCTTCGACTTCAATTTCTTTGGCGAGAAGGC 1020
Q T Q C E G G S T L F D F N F F G E K A 305
CTACCTCACTCAGTCCAGCCAGCTGTATCTGGAGACGTGCATCCCCAGTTTTGGCGATGT 1080
Y L T Q S S Q L Y L E T C I P S F G D V 325
GTTCTGCATTGAGCAGAGTTACCGAGCAGAGCAGTCACGCACCAGAAGGCATCTGGCATC 1140
F C I E Q S Y R A E Q S R T R R H L A S 345
TTACACTCACGTTGAAGCCGAGTGCCCCTTCATTAGCTTCGATGACCTCCTGGATCGCAT 1200
Y T H V E A E C P F I S F D D L L D R I 365
TGAGGATCTGGTGTGCGATGTGGTGGACCGTGTGTTGAAGCACCCCGTGGGCGGAGACCT 1260

```
  E  D  L  V  C  D  V  V  D  R  V  L  K  H  P  V  G  G  D  L     385
GATGAAGGACCTCCACCCTGAGTTTGTGGCTCCCACTCGGCCCTTCCTCAGAATGCCGTA     1320
  M  K  D  L  H  P  E  F  V  A  P  T  R  P  F  L  R  M  P  Y     405
CAAAGACGCCATCCAGTACCTCAAGGACAATGGCATCACCAAAGAGGATGGGACACTGTA     1380
  K  D  A  I  Q  Y  L  K  D  N  G  I  T  K  E  D  G  T  L  Y     425
TGAGTTTGGAGAGGACATTCCCGAGATGCCGGAGCGCAAGATGACAGACAAGATCGGCCG     1440
  E  F  G  E  D  I  P  E  M  P  E  R  K  M  T  D  K  I  G  R     445
ACCAATCCTGCTGAACCGGTTCCCTGCGGGCATCAAGGCCTTCTACATGTCTCGCTGCCA     1500
  P  I  L  L  N  R  F  P  A  G  I  K  A  F  Y  M  S  R  C  Q     465
GGACGACAAGAGCCTCACCGAATCCGTCGACTTGTTAATGCCAGGCGTGGGAGAGATTGT     1560
  D  D  K  S  L  T  E  S  V  D  L  L  M  P  G  V  G  E  I  V     485
CGGAGGCTCAATGAGGATGCACGACTACCAAGAGCTCATGGATGCTTACAAGGCTAATGA     1620
  G  G  S  M  R  M  H  D  Y  Q  E  L  M  D  A  Y  K  A  N  D     505
CTTGGACGCAAAGCCGTACTATTGGTATACCGATCAGCGGCGGTACGGACCAGCCCCCA     1680
  L  D  A  K  P  Y  Y  W  Y  T  D  Q  R  R  Y  G  T  S  P  H     525
CGGCGGCTATGGACTGGGCTTGGAGCGCTTCCTCTGCTGGATCGCCAACCGCTACCACAT     1740
  G  G  Y  G  L  G  L  E  R  F  L  C  W  I  A  N  R  Y  H  I     545
TAGAGAAGCCACTCTCTACCCCAGATTCGTTGGCCGTTGCACGCCTTAAGCAAAGAATGA     1800
  R  E  A  T  L  Y  P  R  F  V  G  R  C  T  P  *                 560
TTAATAAATATATTACAACCACCAGTGTGAAAGAGTTAAAAAAAAAAAAAAAAAA          1835
```

**Figure 3.43** The full length cDNA sequences of *PMATRS*. Start and stop codons are illustrated in boldface and underlined. The 5'RACE-PCR primer is underlined, boldfaced and italicized. The polyA additional signal is boldfaced.

#### 4. *Aspatase amino transferase* (*PMAST*)

A 900 bp fragment was obtained from 5'RACE-PCR of a *PMAST* primer (Figure 3.44). An amplification fragment was cloned and sequenced for both directions. Nucleotide sequences of the original EST and 5'RACE-PCR were assembled and analyzed.

The full length cDNA of *PMAST* was 1944 bp in length where the 5'and 3'UTRs of *PMAST* were 232 and 464 bp (excluding the poly A tail), respectively. The ORF of *PMAST* was 1248 bp encoding a polypeptide of 415 amino acids. The poly A

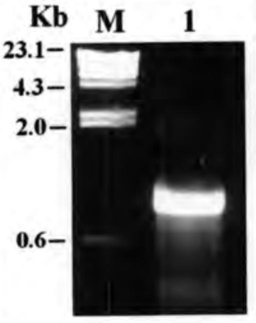
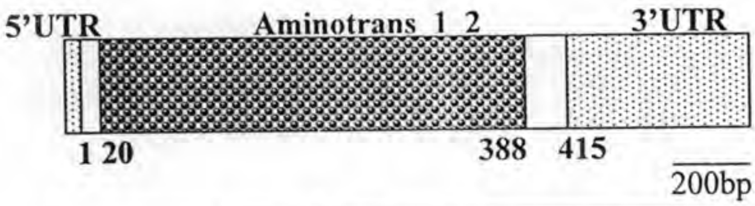



**Figure 3.44** The primary 5'RACE-PCR product of *PMAST* (lane 1). The λ-*Hin*d III (lane M) DNA marker were included as the marker

additional signal (AATAAA) was located between 1904 - 1909 of the entire sequence of *PMAST* (Figure 3.45 and 3.46).

The closest sequence to *PMAST* was *aspartate aminotransferase* of *Tribolium castaneum* (E-value = 5e-176). The calculated pI was 8.55 with the molecular weight of 43293.86 daltons and the signal peptide was not found in this putative nonsecretory protein. An *aminotransferase* domain was found at positions 20 - 388 (8.00e-123; Figure 3.45). *Aspartate aminotransferase* catalyzes the reversible transamination between dicarboxylic amino and a keto acids essentially needs in nitrogen and carbon metabolism in the cells.




| Domain | Position | E-value |
|---|---|---|
| Aminotrans I_II | 20-388 | 8.00e-123 |

**Figure 3.45** Diagram illustrating the full length cDNA of *PMAST.* An aminotransferase class I and II domains were found in this transcript. The scale bar is 200 bp in length.

```
AATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGGAGAGGCCGC    60
CGAGATCTCGCGGTGGGAATACGCTTCGCTCCGCCACAGTCCCGAGGAAATACCTTTCGT   120
CCTTCCCGAAATGGGCCGCGCAGGACCCCGCACGCTCAACCTCGTCAGGCAAAACCAGGC   180
CTCGATCCAGCATGTCTGCTCGAGGGCAAGTTCTTGGTGGTCTGGGGTGGAGATGGACC    240
                                                 M  G  P        3
CCCAGATGCCATCTTGGGTGTTACCGAGGCATTCAAGCGTGACACGAACTCCAAGAAGAT   300
 P  D  A  I  L  G  V  T  E  A  F  K  R  D  T  N  S  K  K  M    23
GAATCTTGGCGTTGGTGCCTACCGCGATGATAACGGCAAGCCTTTCGTCCTCCCGTCTGT   360
 N  L  G  V  G  A  Y  R  D  D  N  G  K  P  F  V  L  P  S  V    43
GAGGAAGGCTGAGGAGCTAATTGTGAGCCAAAAGTTAGACAAGGAGTACTTGCCCATCTC   420
 R  K  A  E  E  L  I  V  S  Q  K  L  D  K  E  Y  L  P  I  S    63
TGGCAGTGCCGAGTTCTGCAAGCATGCTATTACCTTGGCTCTTGGGAGTGACAGCCCAGT   480
 G  S  A  E  F  C  K  H  A  I  T  L  A  L  G  S  D  S  P  V    83
TATTGCCGATGGACTGAATGTAACAGTTCAGGGTATTTCTGGTACTGGCGCTCTCCGTAT   540
 I  A  D  G  L  N  V  T  V  Q  G  I  S  G  T  G  A  L  R  I   103
TGGCTCCACCTTCCTCTCGAAGTTCTTCCCAGGTGCAAAGAATGTATGGCTGCCAGCACC   600
 G  S  T  F  L  S  K  F  F  P  G  A  K  N  V  W  L  P  A  P   123
AACCTGGGGCAACCATGTTCCCATCTTCAAACATGTCAATATGGATGTCAAGCAGTATAG   660
 T  W  G  N  H  V  P  I  F  K  H  V  N  M  D  V  K  Q  Y  R   143
```

ATATTATGACCCAAAGACCTGTGGATTTGACTTCAGTGGAGCAATGGAGGACATTTCTAA 720
Y Y D P K T C G F D F S G A M E D I S K 163
AATCCCTAAGGGTAGTTTGATCATGCTTCACGCATGTGCCCACAACCCCACTGGTGTAGA 780
I P K G S L I M L H A C A H N P T G V D 183
CCCCAAGGCAGAGCAGTGGGACGAAATGAGCAAGGTTATCAAGGAGAGAGAGCTGCTTCC 840
P K A E Q W D E M S K V I K E R E L L P 203
CTTCTTTGACATGGCATATCAA***GGATTTGCCTCGGGAGATGTAGCC***AAGGATGCCTATGC 900
F F D M A Y Q G F A S G D V A K D A Y A 223
TGTGCGCAAGTTCTTGGCTGATGGCCACAAGATCTGTCTTTCCCAGTCTTTCTCCAAGAA 960
V R K F L A D G H K I C L S Q S F S K N 243
TATGGGCTTGTATGGTGAGAGAGCTGGTGCATTTACAATCGTATGCAACGACAAAGATGA 1020
M G L Y G E R A G A F T I V C N D K D E 263
AGCTGCCCGTGTTCTGTCACAGGTGAAGATCTTGATCCGACCCCTTTATTCCAACCCACC 1080
A A R V L S Q V K I L I R P L Y S N P P 283
TCTCCATGGCGCTCGCATTGTGTCCACCATTCTTAGTAATCCAGAACTGAACTCTATTTG 1140
L H G A R I V S T I L S N P E L N S I W 303
GCTGAAGGATGTCAAGGGTATGGCTGACAGGATCATTAACATGCGTACCAAGTTGAAGGA 1200
L K D V K G M A D R I I N M R T K L K E 323
AAACCTGGCCAAGGAAGGGTCCATCAGAGACTGGAGTCACATCACTGACCAAATTGGCAT 1260
N L A K E G S I R D W S H I T D Q I G M 353
GTTCTGCTTCACTGGCATGACTCCAGACCAGGTTGAGAAGCTGACCAAGGAGTTTTCTGT 1320
F C F T G M T P D Q V E K L T K E F S V 373
GTACCTGACAAAGGATGGACGTATCTCAGTTGCTGGTATTGCTTCCAGTAATGTTGAATA 1380
Y L T K D G R I S V A G I A S S N V E Y 393
CTTGGCTCATGCAATGCACCATGCCACCATAAATATAAAAGTTCTCAGATTTAGATCTAT 1440
L A H A M H H A T I N I K V L R F R S I 403
TTATTGGGAGTGTGAAATGGGACTTTTTTACAAT**TAA**TATCACTGAAATGGTTTAGAAT 1500
Y W E C E M G L F Y N * 414
ACAGCCAATGTTAGGACATACACACTATAAACTTTACAAGGTTATTTCTTGTGCTTAACA 1560
TTGCTGTTATCATGATTTTAAAAAGTGCACTGCTTTTCTTAAAATTTAAAAGATTTGCTG 1620
CAGTGTTTATTGATAATCAGTCATGATGAAAATTTGTTCATATGGTAAACAAAGCAGGCA 1680
AAAGATAGTTACCTCCCCCTTTCTGTACATTTCTGTAGATAAGTCAACACCCTACAAAGT 1740
TATTTTAAATTGAACCCAGTGAAATAAAAGACGACCACGTTTGCTTTATTTGAAGATGTA 1800
ATTCATTCTGATATTATTTAAAACCAATTATTTTGGTGTATATACATCCTGTGTCCCTGC 1860
ATTGCAGGTACTATTCCATGTACAGGTACTATTCCATGTACAC**AATAAA**CAGTATAATAT 1920
TTAAAAAAAAAAAAAAAAAAAAAA

**Figure 3.46** The full length cDNA sequences of *PMAST* of *P. monodon*. Start and stop codons are illustrated in boldface and underlined. The 5′RACE-PCR primer is underlined, boldfaced and italicized. The polyA additional signal (AATAAA) is boldfaced.

### 5. *Dolichyl diphosphooligocharide protein glycotransferaase* (*PMDDPG*)

The smear product was obtained from the primary 5′RACE-PCR. After nested PCR was carried out, a 900 bp fragment was obtained (Figure 3.47). An amplification fragment was cloned and sequenced for both directions. Nucleotide sequences of the original EST and 5′RACE-PCR were assembled and analyzed.

The full length cDNA of *PMDDPG* was 2326 bp in length with the 5′ and 3′UTRs of 32 and 992 bp (excluding the poly A tail), respectively. The ORF of *PMADRP* was 1210 bp encoding a polypeptide of 369 amino acids (Figure 3.48 and 3.49).

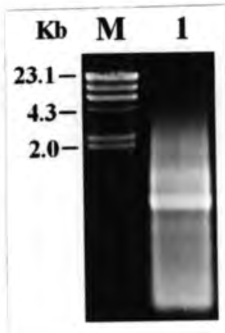
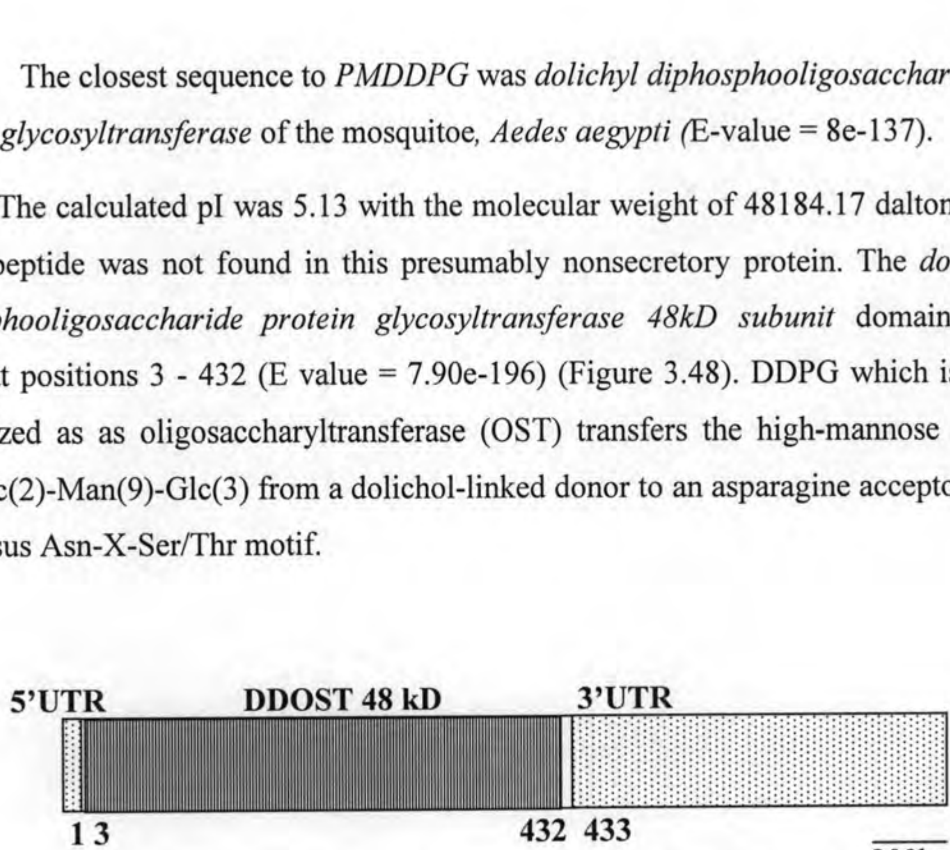



**Figure 3.47** The primary 5′RACE-PCR product of *PMDDPG* (lanes A). The λ-*Hin*d III DNA marker (lane M) were included as the marker

The closest sequence to *PMDDPG* was *dolichyl diphosphooligosaccharide protein glycosyltransferase* of the mosquitoe, *Aedes aegypti (*E-value = 8e-137).

The calculated pI was 5.13 with the molecular weight of 48184.17 dalton. The signal peptide was not found in this presumably nonsecretory protein. The *dolichyl diphosphooligosaccharide protein glycosyltransferase 48kD subunit* domain was found at positions 3 - 432 (E value = 7.90e-196) (Figure 3.48). DDPG which is also recognized as as oligosaccharyltransferase (OST) transfers the high-mannose sugar GlcNAc(2)-Man(9)-Glc(3) from a dolichol-linked donor to an asparagine acceptor in a consensus Asn-X-Ser/Thr motif.




| Domain | Position | E-value |
|---|---|---|
| DDOST_48kD | 3-432 | 7.90e-196 |

**Figure 3.48** The diagram illustrating the full length cDNA of *PMDDPG.* The dolichyldiphosphooligosaccharide protein glycosyltransferase 48kD domain was found in this transcript. The scale bar is 200 bp in length.

```
AAGCAGTGGTATCAACGCAGAGTACGCGGGGGATGATAAGATTAGCAGCATTAGCAGTGC   60
                                 M  I  R  L  A  A  L  A  V     9
TTTTGGCACTGACACTAGCACAGAAACAAAATACTTTGGTGCTAGTGGACACGTTAGCGA   120
L  L  A  L  T  L  A  Q  K  Q  N  T  L  V  L  V  D  T  L  A     29
TTCGGGAGACTCATTCTATCTTCCTGAAGTCACTCCAAGAACGTGGTCATGAAGTTACCG   180
I  R  E  T  H  S  I  F  L  K  S  L  Q  E  R  G  H  E  V  T     49
TGAAAGCTGCTGATGACCCATCACCCCAGCTCTCCAGATTTGGGGAATATATCTATCAAA   240
V  K  A  A  D  D  P  S  P  Q  L  S  R  F  G  E  Y  I  Y  Q     69
ACCTGGTGATTCTTGCTCCAGGAGTAGAAGAATTTGGTGGGGCTCTCAGTGTTGAGGCTA   300
N  L  V  I  L  A  P  G  V  E  E  F  G  G  A  L  S  V  E  A     89
TCGTTGAGTTCATTGATGGCAGTGGGAATGTCCTGGTTGCTGGATCTCGAGAAGCTGCTG   360
I  V  E  F  I  D  G  S  G  N  V  L  V  A  G  S  R  E  A  A     109
ATCTTATCCGTGAGCTTGTGACAGAAGTTGGTGTGGAAATGGATGAGGAAGGAGCAGCCG   420
D  L  I  R  E  L  V  T  E  V  G  V  E  M  D  E  E  G  A  A     129
TCATTGACCATTTGCACTATGATGCAAATGATGATGGGCAGCACACCCTCATTGCTGCAC   480
V  I  D  H  L  H  Y  D  A  N  D  D  G  Q  H  T  L  I  A  A     149
CAAATACTGGACTGATTGACTCGGAGGTTATGGTTGGATCTAATTCCCAAGTGCCATTGT   540
P  N  T  G  L  I  D  S  E  V  M  V  G  S  N  S  Q  V  P  L     169
TGTACCGAGGCACAGGGCTGATCACAGATGCTGACAACCCCTTGGTTCTACCAGTGTTGA   600
L  Y  R  G  T  G  L  I  T  D  A  D  N  P  L  V  L  P  V  L     189
GGGCTCCCTCTACTGCATATTGCTATAATCCTACACAATCCATCACTGACTACCCTCATG   660
R  A  P  S  T  A  Y  C  Y  N  P  T  Q  S  I  T  D  Y  P  H     209
CCACAAGTCAGAATATGCTGNTAGTTGCTGCTCCGCAAGCTCGCAACAATGCAAGAGTGG   720
A  T  S  Q  N  M  L  X  V  A  A  P  Q  A  R  N  N  A  R  V     229
TAGTTTCTGGCTCGCTAGAGTTCTTCTCCGATGCCTTTATCATGGCCTCTGTCCAGACAC   780
V  V  S  G  S  L  E  F  F  S  D  A  F  I  M  A  S  V  Q  T     249
CACAGGGTAAATTTTATGAACGTTCTGGCAACGGCAAAGTAGTAGAAGCTCTTAGTCGCT   840
P  Q  G  K  F  Y  E  R  S  G  N  G  K  V  V  E  A  L  S  R     269
GGGTATTTCGGGAGGAGGGCGTCCTGCGTGTCGTATCTATTGAACATCATCTTCAGGGAG   900
W  V  F  R  E  E  G  V  L  R  V  V  S  I  E  H  H  L  Q  G     289
ATTCTCAACCTCCAGTTGCATACACTATCAAGGAAGATGTGGAGTACAAAATCATGGTTG   960
D  S  Q  P  P  V  A  Y  T  I  K  E  D  V  E  Y  K  I  M  V     309
AGAGGCTTGTGAATGGCTCATGGAAACCATTCATGGCAGATGATGTGCAGATGGACTTTG   1020
E  R  L  V  N  G  S  W  K  P  F  M  A  D  D  V  Q  M  D  F     329
TTCGCATTGACCCATTCATCAGGCTGACCATGACACCCAGTCCTGAAGGGATTTTCTCAG   1080
V  R  I  D  P  F  I  R  L  T  M  T  P  S  P  E  G  I  F  S     349
TCAAGTTTACGGTACCTGATGTCTATGGTGTATACCAGTTCAAGGTGGAGTACAATCGTG   1140
V  K  F  T  V  P  D  V  Y  G  V  Y  Q  F  K  V  E  Y  N  R     369
TTGGTTTTACACGTCTTTACAGTTCCACTCAGGTATCAGTTCGGCCATTCACTCACCGAG   1200
V  G  F  T  R  L  Y  S  S  T  Q  V  S  V  R  P  F  T  H  R     389
AGTACGAGAGATTACTTGAGTGTGCCTTCCCATATTATGCTAGTGCCTTCTCCATGATGT   1260
E  Y  E  R  L  L  E  C  A  F  P  Y  Y  A  S  A  F  S  M  M     409
TTGGTGTCTGGTTATTTCCATGGGTTTTCTTGCACCACAAAGAACCTATCCCCAAACGTA   1320
F  G  V  W  L  F  P  W  V  F  L  H  H  K  E  P  I  P  K  R     429
AGGCCGAATAAGTTTGCTTGTTTGACCTAATAGAACTAGTTAACTTCACCATTCCTGCTG   1380
K  A  E  *                                                     432
TTTCTTTCAGCTGTTGTTTGTGTAGAAGAGATGCACAAGAAGCATATTCATAATTTTACA   1440
TTTTGTAACTACAGGATTTTTCATCATTCTTAATATTGATATACTGTAATTTGGAAACTG   1500
CATGTAAATTATTTTTATTTTTTTTATTTAAATTGGTTAAAGCTGGTTTTCATCACCCTT   1560
AAACCTTCTATGTCTTTTATTTTTTTTTATTTATTTATTTTATTTATTATTATTATTTTT   1620
TTTTTCAATATATGGCTAAAAGGTACATAGTTTAGTCTTTCTTATACATGAAATAATTTG   1680
CATATTTCACCAACAACTTGTAATAACAACTTGTAATTGTTATCTCTGCTAGTAAAACTA   1740
CCTTAAAGGAAAAAACCTCGTGCCGAATTCGGCACGAGGGAGGAGGGTAAGCAGGTTCCT   1800
CAGGAAGATGACTTTGATAGTTTCAAGGAAGGAGTGAAGGGTGTTGTAGTACCTCCTGAC   1860
CAACTGAGTGAATGTGATTTCACAGACGATGAGAGTGGATGCAGTGAAATCCTGGATTAC   1920
AGCCAGATGAGGGAGGATCAGAGCTTCATGCAAGCACAATGGGAAGCCTACAAGAAGGAA   1980
TTTCATCAGAAGCAAAAGGAAGAATCGCAAGAATATCACAAGAAATCTGAGGCAGAAAAA   2040
GAAAAGGATGCTGGTGTACTGGTGAAGAAATGTTTTGAAACATCAGATATAAAGCAAGAG   2100
ATAGCTCTTGGCTTTCCTGAAGGATTTGTGAAAGAGGAGAATATAGATAGTAAGAATCTG   2160
GATTGCAAGATAGAAACGGTTATAGGGAAGAAAGTGTCGAGTGCCAATGATGTAAATACT   2220
GTAAAAAAGGAGGAAAAATGTATGACTACTTTTGAAAAGGATGAAGCAAAAGAAAACACA   2280
GAAACCACTCAGGGTGCTTGTGGAAAGCAAAAAAAAAAAAAAAAAA                 2326
```

**Figure 3.49** The full length cDNA sequences of *DDPG* of *P. monodon*. Start and stop codons are illustrated in boldface and underlined. The overlapped 5'RACE-PCR primer (GCGTCCTGCGTGTCGTATCTATTGAACA) and nested 5'RACE-PCR primer (TCGCTGGGTATTTCGGGAGGAGGGC) are underlined, boldfaced and italicized.

### 6. *Nuclear autoantigenic sperm protein* (*PMNASP*)

A 1620 bp fragment was obtained from 5'RACE-PCR of a *PMNASP* primer (Figure 3.50). An amplification fragment was cloned and sequenced for both directions. Nucleotide sequences of the original EST and 5'RACE-PCR were assembled and analyzed.

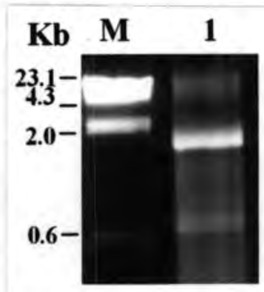
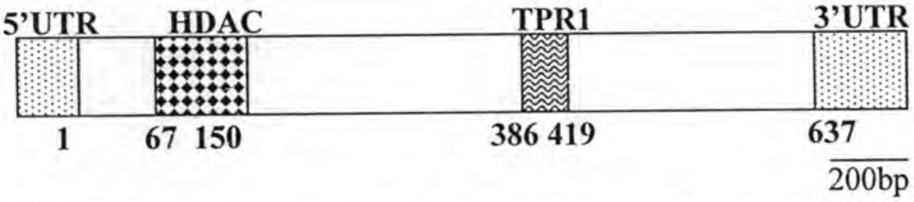



**Figure 3.50** The primary 5'RACE-PCR product of *PMNASP* (lanes 1). The λ-*Hin*d III (lane M) DNA marker was included as the marker.

The full length cDNA of *PMNASP* was 2270 bp in length with the 5' and 3'UTRs of *PMADRP* of 138 and 218 bp (excluding the poly A tail), respectively. The ORF of *PMADRP* was 1914 bp encoding a polypeptide of 637 amino acids (Figure 3.51 and 3.52).

The closest sequence to the full length *PMADRP* was *nuclear autoantigenic sperm protein (histone-binding)* of *Danio rerio* (E-value = 3e-81). The calculated pI was 5.22 with the molecular weight of 70369. The histone deacetylase (HDAC) interacting domain is found at positions 67 – 150 (E-value = 0). The HDAC forms interactions with histone deacetylases and prevent polyspermy during fertilization of eggs and sperm.

The tetratrico peptide repeat (TPR) was also found with low matching value. TPR is a structural motif present in a wide range of proteins It mediates protein-protein interactions and the assembly of multiprotein complexes Proteins containing TPRs are involved in a variety of biological processes, such as cell cycle regulation, transcriptional control, mitochondrial and peroxisomal protein transport, neurogenesis and protein folding.




| Domain | Position | E-value |
|---|---|---|
| HDAC_interact | 67-150 | 0.00e+00 |
| TPR_1 | 386-419 | 7.70e-02 |

**Figure 3.51** Diagram illustrating the full length cDNA of *PMNASP*. The HDAC interacting domain and tetratricopeptide repeat domain were found in this transcript. The scale bar is 200 bp in length.

```
CTAATACGACTCACTATAGGGCAAGCAGTGGATCAACGCAGAGTACGCGGGGACTCTTTT   60
CCGGCCTTTGTACGCTCCACACGCGCAACCCCGTCACCCTCGTTTGTTGCTTCCGAACGT   120
TGCTAATAGTCGACCACGATGTCCGAGTCGCCGGTAAAAGCTGCCGAGACCAGTGCCAGC   180
                  M  S  E  S  P  V  K  A  A  E  T  S  A  S     14
CCCAAGAGCTCGCCGTCCAAGAAAGAGATTGATACAGCAACCCAAGCTTTAAATCACTTT   240
 P  K  S  S  P  S  K  K  E  I  D  T  A  T  Q  A  L  N  H  F    34
GCTCAGGGCAAGAGACACTTGGTTGTTGGTGACATTTCATCTGCAGTTAATTCTTTGCAG   300
 A  Q  G  K  R  H  L  V  V  G  D  I  S  S  A  V  N  S  L  Q    54
GAAGCATGTAGACTACTAGCAGAGCAATACGGTGAAACTGCTCCCGAGTGTGGTGATGCT   360
 E  A  C  R  L  L  A  E  Q  Y  G  E  T  A  P  E  C  G  D  A    74
TATTTCTACTATGGCCGTGCACTGCTTGAAATGGCACGCATGGAGAACGGAGTCTTAGGA   720
 Y  F  Y  Y  G  R  A  L  L  E  M  A  R  M  E  N  G  V  L  G    94
AATGCTTTGGATGGAGTTCCCGATGGAGAGGACATGGACAATTCCCAGGTAGAAAATCCT   780
 N  A  L  D  G  V  P  D  G  E  D  M  D  N  S  Q  V  E  N  P    114
GAAAAAATGACAGAGGATGAGAAGAACGAGGTAACAGAACAGGTTGGGAAGGCATTGGAA   840
 E  K  M  T  E  D  E  K  N  E  V  T  E  Q  V  G  K  A  L  E    134
GAGAACTTTAAAGATCTTGAGGATGTGTCAAAAAGTAAGTCGGCACAGCAGAATGGAGAT   900
 E  N  F  K  D  L  E  D  V  S  K  S  K  S  A  Q  Q  N  G  D    154
GCAAAGGCAAAGGCAGAAGAGTCTTCAGGTGTTGAGGAGGCTAANATGGATGTAGATTCA   960
 A  K  A  K  A  E  E  S  S  G  V  E  E  A  X  M  D  V  D  S    174
GCTGGAGTGTCAGAATCAAAGGTGAAGATGGAGGCGAGAGGCAAGAGGATAAGTAGAAAG   720
 A  G  V  S  E  S  K  V  K  M  E  A  R  G  K  R  I  S  R  K    194
TCGGAAGGGGAGGAAAAGAGTAAAGAGGAAACCTCGGACACTGATGGCACCACCACTTCC   780
 S  E  G  E  E  K  S  K  E  E  T  S  D  T  D  G  T  T  T  S    214
AAAGTAGAGGCTAGCTCAGTAGATAGTGAAAAGGTAGACAAGGAAATCAAGCCTGAGAAA   840
 K  V  E  A  S  S  V  D  S  E  K  V  D  K  E  I  K  P  E  K    234
AAGGAAGTTGTGGATACCAAAGATAGTTCCAAAGAGGAGGCAGAGGAATCCGAAAAGGTG   900
 K  E  V  V  D  T  K  D  S  S  K  E  E  A  E  E  S  E  K  V    254
ACGGAGGAGAAGGTTGAGGCTAAGGAGGAAGAAGGGAAAACCACTGAGAAGGGAGAGGGA   960
 T  E  E  K  V  E  A  K  E  E  E  G  K  T  T  E  K  G  E  G    274
GAGAAGGAAAAGGGATCAGGAGACACTAAAGATGAAAAGGGAAAGGAAGATGCCAAAGTG   1020
 E  K  E  K  G  S  G  D  T  K  D  E  K  G  K  E  D  A  K  V    294
GAAGAGGAGAATGTAAAAACTGAAGCAAAGGAAGAGAAATGGAAACTGATTCGTTGTGAG   1080
 E  E  E  N  V  K  T  E  A  K  E  E  K  W  K  L  I  R  C  E    314
AAGGAAGGCAGTTCCAATGGAGAGAGGAAGAGGAAGCGGATGGCGATGGTGATGGTGATG   1140
```

K E G S S N G E R K R K R M A M V M V M 334
GCGAAGGCGAGGGAGAAGATTCTCAAGAAGATTCCCAGGATGAAGGTGAAAAGGAAAGTA 1200
A K A R E K I L K K I P R M K V K R K V 354
TCAAACTTGCAGCTCTCTTGGGAGATGTTGGAATTGGCAAAGGTTATCTATCAGAAGCAA 1260
S N L Q L S W E M L E L A K V I Y Q K Q 374
CAGGATGACAACCCAGAGATGGCTAAGAAAGTTGCCCAAGTGTACCTAAAACTTGGAGAA 1320
Q D D N P E M A K K V A Q V Y L K L G E 394
GTAGGCTTGGAGAGCGAAAATTATTCACAGGGTATTGAAGATTTCAAACAGTGTCTGCAA 1380
V G L E S E N Y S Q G I E D F K Q C L Q 414
ATACAGGAGAAAATTCTTGAGGAAGACAACAGGTGTTTGGCAGAAACCCATTACCAGCTT 1440
I Q E K I L E E D N R C L A E T H Y Q L 434
GGTGTAGCACACTCCTTCTGCTCGAAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGC 1500
G V A H S F C S K L T L T K G N K S W S 454
TCCACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCCGGGCTGCAGGAATTCGGCA 1560
S T A V A A A L E L V D P P G C R N S A 474
CGAGGAGAAGAGCAAAGAAAGAAAGATGCTGCAGAGAGACCTGATCCATTCTACACCGAA 1620
R G E E Q R K K D A A E R P D P F Y T E 494
GAAGGCGAGATTGAAGAATTGAACAAATTGTTACCAGAGATGAAGGAAAAGGTTACAGAT 1680
E G E I E E L N K L L P E M K E K V T D 514
ATGGAGGAAATGAAGAAAGA***CAGCAAGGACAGACTCCAGAAAGCGGC***AAAGGAAGCATTC 1740
M E E M K K D S K D R L Q K A A K E A F 534
ATGGCAAATGCAATTGGTGGCACCTCCAAAGCTGGATCTTCATCACAAACTGGATTTGAT 1800
M A N A I G G T S K A G S S S Q T G F D 554
GCACCTTCAAGTTCTACATCCTCAACCCCCACAGAAATAAAGGCTTCCAACATTACTCAT 1860
A P S S S T S S T P T E I K A S N I T H 574
CTTGTAAGAAAGAAGCAGAGGAAACCTGAGGATGAGGTTGAGGGAGAAGAGGTGAAAAAG 1920
L V R K K Q R K P E D E V E G E E V K K 594
GCAAAAGGCGAGAATGGGGAACCACATGGAACTGCTAATGGACCACCAATGGCACCAATG 1980
A K G E N G E P H G T A N G P P M A P M 614
GGCACTCTGAAACCATGGAAACAGAGGAAAAGGATACCCCAACAAATGGGGCAAGCACTG 2040
G T L K P W K Q R K R I P Q Q M G Q A L 634
AAGAAT**TAA**AGGAGAAAGCAGCTGAAGAGATGAAGAAAAAGACGGATATGATCACTGGGA 2100
K N * 636
AAACTGAGGCAGCATCCTAGATAGTGTATAATGCCTTTAAAACACTGCCAGTTATTGGAC 2160
TAAGAACTTTTGTATTTTATTGTAATGGAAGCCATTTTATTACCAGTGAACTGTGTTTTA 2220
AGATTTACATAATTTATTAAAAATGAAATGTAAAAAAAAAAAAAAAAAA 2270

**Figure 3.52** The full length cDNA sequences of *NASP* of *P. monodon.* Start and stop codons are illustrated in boldface and underlined. The 5'RACE-PCR primer are underlined, boldfaced and italicized whereas the internal sequencing primer is boldfaced.

### 7. *Female sterile* (*PMFS*)

A fragment of approximately 3000 bp long was obtained from 5'RACE-PCR of a *PMFS* primer (Figure 3.53A). An amplification fragment was cloned and sequenced for both directions. Primer walking was carried out to obtain the adjacent sequence of the 5'RACE-PCR product but it was surprisingly not successful. The 5'FSII forward primer (5'-TGA TAG CCT GGA GGA TGA-3') and 5'FSII reverse primer (5'-GAC AGG CTC CCA AAC CAT-3') and the 3'FSII forward (5'-ACA ACT TGG TGG TGC TCT CG-3') and 3' FSII reverse (5'-CCA CAT TAG TAG CCA TAA CAT C-3') was designed and successfully amplified the fragments of 1300 and 1500 bp, respectively.(Figuer 3.53B) The amplified fragment was cloned and sequenced. Nucleotide sequences of the original EST, 5′RACE-PCR and RT-PCR were assembled and analyzed.

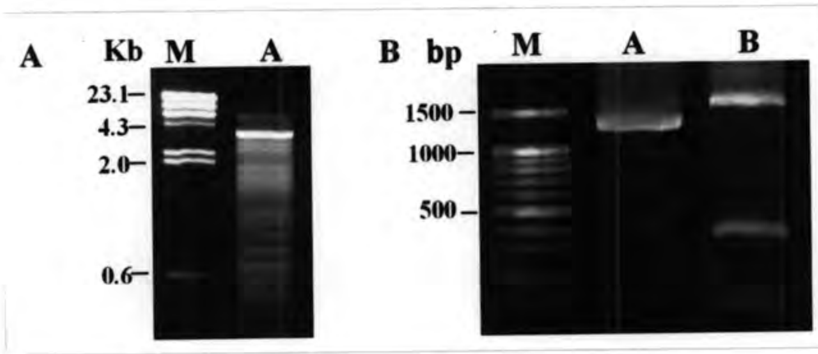
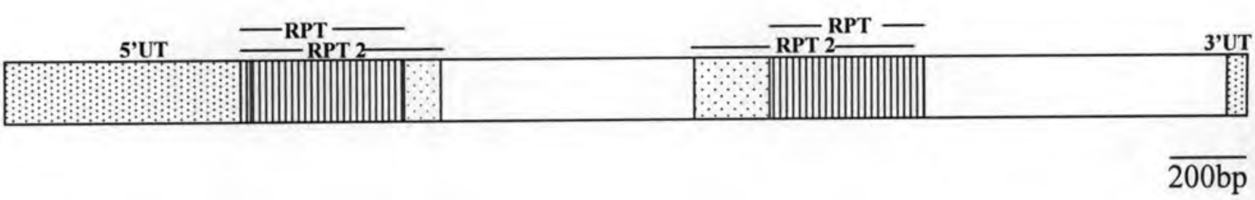



**Figure 3.53** The primary 5′RACE-PCR of *PMFS* (panel A, lane 1) and the conventional PCR product of 3′ (lane 1, B) and 5′ fragment (lane 2, B) of *FS.* A λ–*Hin*d III (lane M, A) or100 bp DNA ladder (lane M, B) were used as the markers

The full length cDNA of *PMFS* was 5028 bp in length with the 5′ and 3′UTRs of 1191 and 70 bp (excluding the poly A tail), respectively. The ORF of *PMFS* was 3945 bp encoding a polypeptide of 1314 amino acids. (Figure 3.54 and 3.55).




| Domain | Position | E-value |
|---|---|---|
| RPT_2 | 3-267 | 8.37e-07 |
| RPT_1 | 6-201 | 1.13e-07 |
| RPT_2 | 590-866 | 8.37e-07 |
| RPT_1 | 687-884 | 1.13e-07 |

**Figure 3.54** Diagram illustrating the full length cDNA of *PMFS*. Internal repeat 1 and 2 domains were found in this transcript. The scale bar is 200 bp in length.

The closest sequence with the full length *PMFS* was *ficolin* of the sea urchin, *Strongylocentrotus purpuratus* (E-value = 7e-37). The calculated pI was 4.46 with the molecular weight of 141312.27 dalton and the signal peptide was not found in this putative nonsecretory protein. Four internal repeat domains were found in *PMFS* (Figure 3.54).

```
CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGTACGCGGGATGTGTAT   60
GGTTATTTCAAGCCCAGCTTTGTAATGCAAACGCACTGCCACTCCATACATAATCTTCTC   120
ATGAAGAGACTAAGCAATACTTATGTGATTGTGTATGTCTGTAAGGAAGGTGTAAAAAGC   180
AGTCTCGACTCTAGGGAAGTGATCATGGACGAAGTAAATCTAGAGAAACCAAAAGACAGG   240
GAAGATCTTCTTCTCGAGTGTCTTGTCGACTTACAAGCAGACTTGGATGGCAGAAAACCA   300
ACAATTTCTCTGCTTGAACAAGCAATTGCAGATGAAATTGTGATGACTAATGATGTATCA   360
CAGACCTGGCAAGAAACTCAAATCTTCACAGCTGGTCTGACCATTGGTGGCACAACAGGT   420
ATACTGCAGACTGCAAATGTAAAGGGTCAAGGTACAACACAACCACATGAAGAAACTTTC   480
GAAGAGTTCTACGCAAAGACTGATAGCCTGGAGGATGATGTAGAGGCCATTGGTGCTAGC   540
TTGGATGGTGTCTTGTATCATTCAACAAACCAAGTGCTTTCAGGGTCTGTTGTTGCTTCT   600
GTCATAACTGCAGACAAATTCGAATCGAATAGCACACGGGTTACAAAGTTAAATGATATG   660
CCGGTGCTGGATATGTCTAGCACATTTATGATTGATGGAATAGAGCAGAACATAAGGTCT   720
TCAATGTACGTTCATTCAATGAAAACGGATCTCTTCTCTACTAGAGAGCAAGGCTTGTCT   780
GCACCTATTAATGATGTGCTTACCAACCAATTCATGAGGAAGTCTGTTGCTTCACAAGAA   840
GTAACGGGTAATCATCACTACAGTGATTTCCATGTAAACCACATCCATGGCAAGCAGGGT   900
GATGGATCTCCATTAGTCCTTAATGGAATAGACACTAGTAAAATTGTGACTAAAGGAAAC   960
AATTTTACCTTTATGGACAAGAAAACCTTCCATAGCATGACCATTCTGCAAGAGTTAGAT   1020
GTTCAACTTCTCAATGAAGTCGACCTATCTGATCTTTCAAGCCATTTGGTGTACACTAAT   1080
GTGTACAGGCCCCAAATACTGGATGGCTCATTTACTCTACATGATGTAAATGTGGATGGC   1140
AACGTTGATGTAAAAAACATCAATAGTGTCGACATGAAAGAACTTAACACCATGTCGTC    1200
                                                   M  V  V     3
AGAACTTCTGGAGACTTTACATTATCTGGTGATGTAACCTACCAGAAAGACTTCCATGTG   1260
 R  T  S  G  D  F  T  L  S  G  D  V  T  Y  Q  K  D  F  H  V    23
TCAGGAAACCTGATCAGCCCAATACTAAATGGTATTGTTATGGATAACATTGTAGATAAA   1320
 S  G  N  L  I  S  P  I  L  N  G  I  V  M  D  N  I  V  D  K    43
GATACCACCACCATAAATGGTGTCTATACCTTTACAAATGCAAACATTAAAGCAGCTATA   1380
 D  T  T  T  I  N  G  V  Y  T  F  T  N  A  N  I  K  A  A  I    63
GGCTGCTCCAACATCAGTGGAATAAATCTAAGCGTAGATGTTGTGACTGTTGATGCTGAC   1440
 G  Ç  S  N  I  S  G  I  N  L  S  V  D  V  V  T  V  D  A  D    83
CAGACTATATCGGGCGCTTTAACCTTCACTGACGACGTGTTGGTAACTGGCCCTGAAGGA   1500
 Q  T  I  S  G  A  L  T  F  T  D  D  V  L  V  T  G  P  E  G    103
GTAAAGATGTTGGATTCTGTTACCATTAATAACATCGACCCCTATAGCCTTGATAAGATG   1560
 V  K  M  L  D  S  V  T  I  N  N  I  D  P  Y  S  L  D  K  M    123
GATGACCATGGAAACCTCTTCGTAGAAAAGGCTGTTGTCTTTAATGCACCACTTCATGTG   1620
 D  D  H  G  N  L  F  V  E  K  A  V  V  F  N  A  P  L  H  V    143
ACAGAGGATGTAGATGTTGAAGTTATCAATGCATTGGCACTCAAAGGCATAGAAGACCGT   1680
 T  E  D  V  D  V  E  V  I  N  A  L  A  L  K  G  I  E  D  R    163
TATTGGAGAAAGGAAACTGATCAAGTAATTGATGTCTTGCCAGAGATTGTATCTACCACT   1740
 Y  W  R  K  E  T  D  Q  V  I  D  V  L  P  E  I  V  S  T  T    183
TTCAGTGATTATGTGACTGCTAAGAATATAAACAACCACCAGATGGCAGACTTTTTGTCT   1800
 F  S  D  Y  V  T  A  K  N  I  N  N  H  Q  M  A  D  F  L  S    203
GTGACAGGCTCCCAAACCATCAATGGAGCCTATACCTTCCAGGGATTGGTAACCATAAAT   1860
 V  T  G  S  Q  T  I  N  G  A  Y  T  F  Q  G  L  V  T  I  N    223
GGACATCTCAAAGTAACAGATGGCAAAGTAATAGATGGTGTGGATGTATCTTCACTACAT   1920
 G  H  L  K  V  T  D  G  K  V  I  D  G  V  D  V  S  S  L  H    243
GATAACTTGGTAACTCTATCTGATAATCAAGACATAGAAGCAGAGACCACCTTTGGTAAG   1980
 D  N  L  V  T  L  S  D  N  Q  D  I  E  A  E  T  T  F  G  K    263
GTCATCATATTGGGTGACTTGGTGCTGAATGGTGACCTCAATGGATGGAATGTTGTGGCC   2040
 V  I  I  L  G  D  L  V  L  N  G  D  L  N  G  W  N  V  V  A    283
GACTTGGTACGGCTTGACCAATCCCTGCCCCAAACTGGGAGTCTTGCATTTTTGGACAAA   2100
 D  L  V  R  L  D  Q  S  L  P  Q  T  G  S  L  A  F  L  D  K    303
GCTACAGCAACATCTCTGCAGCTGTCTAGTGCAGACCTTACTGTTCAGAGCCTTAATGGA   2160
 A  T  A  T  S  L  Q  L  S  S  A  D  L  T  V  Q  S  L  N  G    323
```

ATGGATGTGAAATCTGCTACAGAGGATTTGGTCTTGGTAAATGAAGATGCATCACTTGCT 2220
M D V K S A T E D L V L V N E D A S L A 343
GGGCCCCTGAAGTTCACCTCCAATACGAAAGCTAATGACCTCTTCGTTAGTGGAACTGTT 2280
G P L K F T S N T K A N D L F V S G T V 363
GATGGTGTTGATGTGACGGATCTTGTAGACCGCAGCCTTAAGAAGACTTCTGCTACACCA 2340
D G V D V T D L V D R S L K K T S A T P 383
CAGGCAGTAACGGGG***GCAATAACGGTGAACAAGGGAG***TTCACTTTGATCAGAGCCCATCT 2400
Q A V T G A I T V N K G V H F D Q S P S 403
TTGACCATGGTTAACAGCAAGGACTGGACCACCTACCTTAGCAAGGTTGTGCCACAAAAT 2460
L T M V N S K D W T T Y L S K V V P Q N 423
TACAATGGTGCAATTGGCGGAAAGAAGACTTTCACAAAGCCAGTATCTATATCCGGCAAC 2520
Y N G A I G G K K T F T K P V S I S G N 443
TTCAACCCAACTACACTAAACGGGTTTAGTGTAGTTCCACTATCGGACAGAATACTGACA 2580
F N P T T L N G F S V V P L S D R I L T 463
AAGAGCACAAACCAGAACGTTGGCAGCAAGTACACCATCAATGGG*GATGTTATGGCTACT* 2640
K S T N Q N V G S K Y T I N G D V M A T 483
*AATGTGG*TTGCAGCAGAAATTGATGGAGTGTTGTCCTCAAATCTCCTCCTCCTAGATGAG 2700
N V V A A E I D G V L S S N L L L L D E 503
AGCAGTATTGTATCTGGCATGGTGGACTTTGCTGATAACTTAATCATTGCTGATGTAACG 2760
S S I V S G M V D F A D N L I I A D V T 523
TCCGAGTCTCGAGTTCTTGATGGATGCAATGTAGTTCAGTTAAACACCTCAACCATCTGG 2820
S E S R V L D G C N V V Q L N T S T I W 543
AAAAATGGTAATGGAGATGTGGTAATGCCCTTCAACATGGCAGTAACAAATCTTCTAGTT 2880
K N G N G D V V M P F N M A V T N L L V 563
AAAAAAGATGCAACTGCAAAGGGTCCAGTAAAAGCTGGAACAAGCCACATGGATGTCTTC 2940
K K D A T A K G P V K A G T S H M D V F 583
CATTTCCTGGATAAGATAGTTACAAAGTCTTCCAACCAAGAAATAACAGGCACAGTAGAG 3000
H F L D K I V T K S S N Q E I T G T V E 603
TTCATGACAAACTTGTCTGTGAATGATCTTCTAACAAACACCATAGATGACGTTTATGTA 3060
F M T N L S V N D L L T N T I D D V Y V 623
GACAATCTCTATGCTGTGACTGTTATGGACAATGAAGCAAGTGTAATAGACTGTGATACA 3120
D N L Y A V T V M D N E A S V I D C D T 643
GACTTCACCAAAGTTCTGACGGTTGATAATCTGAAAGTAAAAACTTCCCTGCACGGATCT 3180
D F T K V L T V D N L K V K T S L H G S 663
GGTGTTGAAGGGGTGTTGATAAATACAATGAATGTAACGGATGTGAATACCCATGCAGTC 3240
G V E G V L I N T M N V T D V N T H A V 683
CACTTAACGGGTGGACCATACATGATAACTGGTGACAAAACTTTCAACAGTGGTCTCTCT 3300
H L T G G P Y M I T G D K T F N S G L S 703
GTTGGCGAACTTGCTATTGATGGATCACTAGATGGAGTACCTGTAG*ACAACTTGGTGGTG* 3360
V G E L A I D G S L D G V P V D N L V V 723
*CTCTCG*GATCATAATAGACATGCTGCCAATATCCTCTTCAAAGCTCCCATCTCAATTAGT 3420
L S D H N R H A A N I L F K A P I S I S 743
GGAGATCTACAGGTTGATGGGCTGTTGGATAATGTGAACCTTGAGCAGCTCCTATCTGAC 3480
G D L Q V D G L L D N V N L E Q L L S D 763
AGGATCAAACTAGATGCTACAGAAACATTAAGCTCATCTACTATATTTGATGGAGTGAAA 3540
R I K L D A T E T L S S S T I F D G V K 783
GTGGAAGGTGACCTCTATGTTGATACCATAGACGGAATCATGGTGTCAGAGATTGTCTTT 3600
V E G D L Y V D T I D G I M V S E I V F 803
AAATCGGGAAGGATGCAGCAGGAGATAGAAGGAGTAAAGACCTTCTCTGGAGGGTTACAT 3660
K S G R M Q Q E I E G V K T F S G G L H 823
GTAGTGGGTGAAACGCAGGCCCCTGTGGTAAATGGAATAAATATTCTGGATTTGAATAAT 3720
V V G E T Q A P V V N G I N I L D L N N 843
AATGTGGTCCGGAAAGACAGGGCAGCAACTATAACTAAAGAGTTGGTCTTCGAGAAGCCA 3780
N V V R K D R A A T I T K E L V F E K P 863
ACAATATCGCAAGTAGATATGTTGGTACAAGGAAATGTAAATGGTTATGATCTTTCTGAA 3840
T I S Q V D M L V Q G N V N G Y D L S E 903
ACCGACTATGAGGCCTCGGTCCTACAAGGAAACATTAAAGCAGAGAATGACCGCTTATTG 3900
T D Y E A S V L Q G N I K A E N D R L L 923
AACCTAAACTTGACTTTGTCAACTATCCATGTTGACACAAAGCTACTCTCCTGTGGAATG 3960
N L N L T L S T I H V D T K L L S C G M 943
TATGAAACGTACCATTATGGTGAAAGAATAAACGAAAAGGCTATATCCATTTCTGGTAAA 4020
Y E T Y H Y G E R I N E K A I S I S G K 963
ATGAGTTGTGGAACATTCGGAGGTTCATCAGTATTGGCTGTGCGGGATTGCAGTGACTAC 4080
M S C G T F G G S S V L A V R D C S D Y 983
GATTGCAGATGCCCTTTGCAATATGCACTCTATGAAGTTGATGATAACGGTAACATGAGC 4140
D C R C P L Q Y A L Y E V D D N G N M S 1003
CAGATAGAGACGGATGTAAACTCTGCTTTCATCTTCAGTACAGAAGGGTATGAAGGCACT 4200
Q I E T D V N S A F I F S T E G Y E G T 1023

GGACTAATAAGCAGCTGTGCCAATGGTGGATCAAGTACTGTAAAGATCTTGCTAAACAAT 4260
G L I S S C A N G G S S T V K I L L N N 1043
AGGGAAACGGACCTGCCACAGGGATCAGCACTAGGAATTATTGCAGATGCAAAATCCTTT 4320
R E T D L P Q G S A L G I I A D A K S F 1063
ACCACCAGTGGTGGCACCTACATGGTGACGGCAGGTGCAATTTCTGATGTGAATACTGCT 4380
T T S G G T Y M V T A G A I S D V N T A 1083
CCAACAACAAATACCACCAGTGGTGGCACCTACATGGTGACGGCAGGTGCAATTTCTGAT 4440
P T T N T T S G G T Y M V T A G A I S D 1103
GTGAATACTGCTCCAACAACAAAAGTCAGTGTGTTAAAGTTGAATAACAATGCCATAGAT 4500
V N T A P T T K V S V L K L N N N A I D 1123
GTGATTTGGTCTCTAGATACTGCCTATAGTGCTTCAACACTCGACTTGACCCTGGGAGAT 4560
V I W S L D T A Y S A S T L D L T L G D 1143
GAAGGTTGGTTGCTCTTGGTAGCAAATTTGATGGCAGCAAATGATACTGTAGATCCTTTC 4620
E G W L L L V A N L M A A N D T V D P F 1163
ATGGCCCCATCACAATTGTATCTCTGGTCTACTGCAGAAGAAAAGTTCACCCTTGTTCAA 4680
M A P S Q L Y L W S T A E E K F T L V Q 1183
GAATTTATGGGTCAACATGTGACCTCGGGCATCTTTTTGAATTCTAAGAAGAATCTCAAG 4740
E F M G Q H V T S G I F L N S K K N L K 1203
GAAAGATTCTTTACTCTTACTCAGTACAAAGCGGCAAAATCTCAAATGCAGGGAACAAAA 4800
E R F F T L T Q Y K A A K S Q M Q G T K 1223
TTAAATACAAAAGTGCAGGTGTTCAAGTACATAGATTCTCGTTATGTTCCATTCGAGAGT 4860
L N T K V Q V F K Y I D S R Y V P F E S 1243
TTACCAACGCTTGGAGCTATTGCCCAAGCCTCGCTTTCCATAGGCGATGACCTCTACTTG 4920
L P T L G A I A Q A S L S I G D D L Y L 1263
GCAGTACTTTCGGACTTTAAAAGTACCCTAGATATCTATGAACTCTTACCTTTTGAGGGA 4980
A V L S D F K S T L D I Y E L L P F E G 1283
TTCCATTTGCAGCAGAGCATTGCAGTATGTGAGACTCCCCTTGATGTGAAGAAGATAGAG 5040
F H L Q Q S I A V C E T P L D V K K I E 1303
ACGACAGAGGACTTGCTTATTGTTACTTGTACAAATCCAAATCAAATGATAACAATAAGA 5100
T T E D L L I V T C T N P N Q M I T I R 1323
TTAAATGCCAAAGGATATTCTTACAGATTA**TAA**TGGAGTGTGGAGCTTGTTTATTTTAAC 5140
L N A K G Y S Y R L * 1333
TTTTAAATACAAATCTATGGTAAAAAAAAAAAAAAAAAAAAAAAA 5185

**Figure 3.55** The full length cDNA sequences of *FS* of *P. monodon*. Start and stop codons are illustrated in boldface and underlined. The 5′RACE-PCR primer are underlined, boldfaced and italicized whereas a primer for primer walking is boldfaced. Primer for cloning of 5′ PCR fragment (5′-TGA TAG CCT GGA GGA TGA-3′, in boldface and italicized) was used in combination of the 5′RACE-PCR primer. Primer for cloning of the 3′ PCR fragment are underlined and italicized.

### 8. *Endothelial cell growth factor I* (*PMECGFI*)

Fragments of 750 and 1700 bp was obtained from 5′ and 3′RACE-PCR of *PMECGFI* primers (Figure 3.56). An amplification fragment was cloned and sequenced for both directions. Nucleotide sequences of the original EST and 5′ and 3′ RACE-PCR were assembled and analyzed.

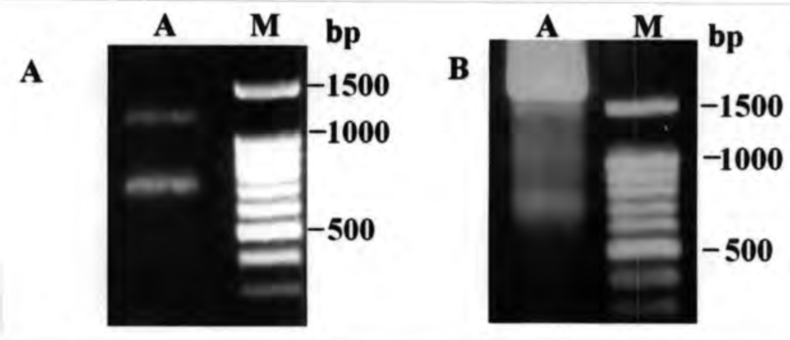
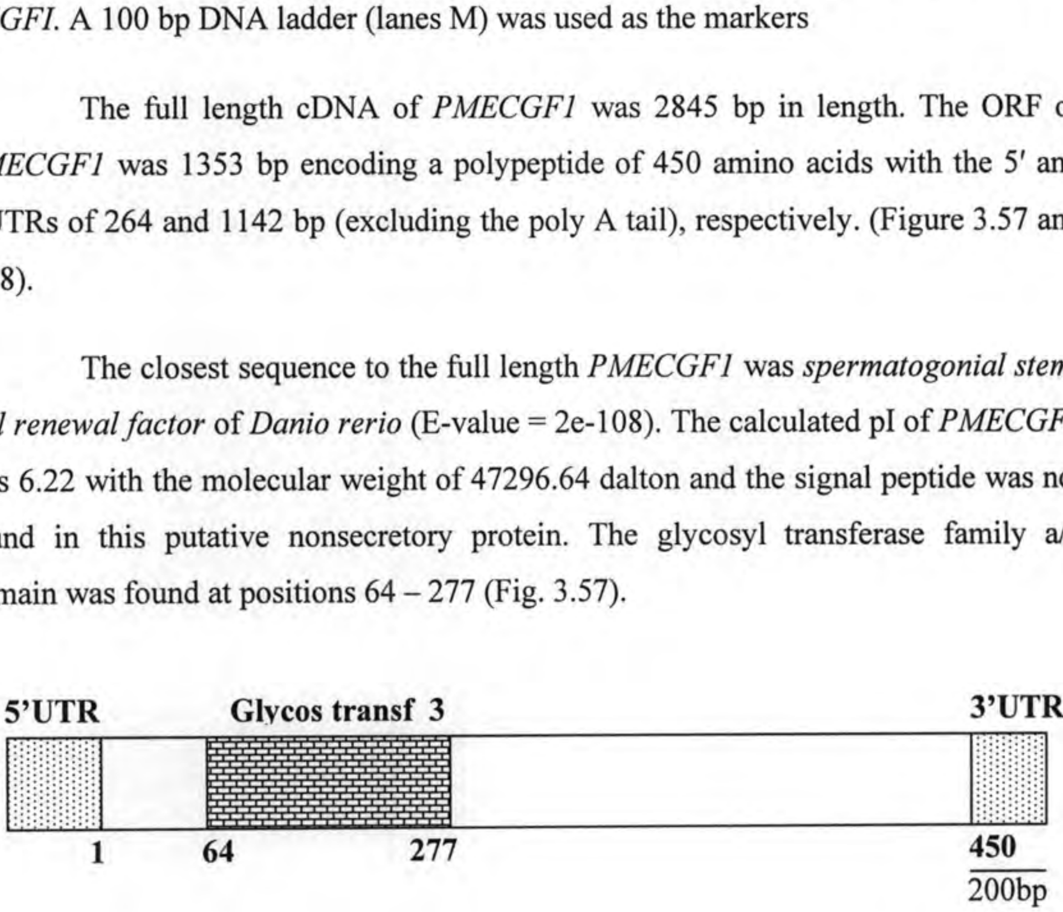



**Figure 3.56** The primary 5′ (lane 1, A) and 3′RACE-PCR (lane 1, B) product of *ECGFI*. A 100 bp DNA ladder (lanes M) was used as the markers

The full length cDNA of *PMECGF1* was 2845 bp in length. The ORF of *PMECGF1* was 1353 bp encoding a polypeptide of 450 amino acids with the 5′ and 3′UTRs of 264 and 1142 bp (excluding the poly A tail), respectively. (Figure 3.57 and 3.58).

The closest sequence to the full length *PMECGF1* was *spermatogonial stem-cell renewal factor* of *Danio rerio* (E-value = 2e-108). The calculated pI of *PMECGF1* was 6.22 with the molecular weight of 47296.64 dalton and the signal peptide was not found in this putative nonsecretory protein. The glycosyl transferase family a/b domain was found at positions 64 – 277 (Fig. 3.57).




| Domain | Position | E-value |
|---|---|---|
| Glycos_transf_3 | 64-277 | 6.20e-02 |

**Figure 3.57** Diagram illustrating the full length cDNA of *ECGF I*. The Glycosyl transferase family, a/b domain was found in this transcript. The scale bar is 200 bp in length.

The glycosyl transferase family is composed of anthranilate phosphoribosyltransferase (TrpD) and thymidine phosphorylase. These proteins transfer a phosphorylated ribose substrate. Thymidine phosphorylase catalyses the reversible phosphorolysis of thymidine, deoxyuridine and their analogues to their respective bases and 2-deoxyribose - 1-phosphate. This enzyme regulates the availability of thymidine and is therefore essential to nucleic acid metabolism.

```
GGACGAGACGATGAGCGGCACAGATTTGACGGTTGAAGCGACGTCTCTTGCGGCGTACAT   60
GGGATTCATGGGTGTGTGCGTGTGTGCGTTTTCGAGAGACCACTAGTGGGAAAGCCTCGC   120
GGAGAAGTGCAAGTTCAGTCGCGGCGGAGTTTCGCAGCGGGACCGGTTGGTGTGCTCCTC   180
GTGGTGGCCGCACTGTTATCTACACCTGGTCCCCACCCGAGTACGATTCGCCGCACAACC   240
GGTCAGATAGGGGCCACGGCCACCATGAGCGCTGCACAGACCTCTCAAGGCAGGTGGAGG   300
                        M  S  A  A  Q  T  S  Q  G  R  W  R     12
ATCCCGGACCTTCTGTCCATGAAGCGCGATGGCCTGGCCTATTCGGAGGACCAGATCGCC   360
I  P  D  L  L  S  M  K  R  D  G  L  A  Y  S  E  D  Q  I  A     32
TTCTTGGTCCGGTCGGTCTCGGATCGGTCCATGGACGACTGTCAGCTGGGGGCGCTCCTG   420
F  L  V  R  S  V  S  D  R  S  M  D  D  C  Q  L  G  A  L  L     52
ATGGCCATCAAGCTGCAGGATATGACGGACGTAGAGACGATCGCCCTCACTAAGGGCATG   480
M  A  I  K  L  Q  D  M  T  D  V  E  T  I  A  L  T  K  G  M     72
AGGGACTCAGGAAGTGTGTTCTCGTGGCCGAAGGACTGGCGCGTCGTGGACAAGCACAGC   540
R  D  S  G  S  V  F  S  W  P  K  D  W  R  V  V  D  K  H  S     92
ACGGGCGGCGTGGGTGACAAGGTGTCCCTGGCGCTGGCCCCCGCCCTCGCCGCCTGCGGC   600
T  G  G  V  G  D  K  V  S  L  A  L  A  P  A  L  A  A  C  G     112
TTCAAGGTGCCCATGATCAGCGGGCGCGGCCTCGAGCACACGGGGGGCACGCTCGACAAG   660
F  K  V  P  M  I  S  G  R  G  L  E  H  T  G  G  T  L  D  K     132
CTGGAGAGCATCCCAGGCTTCAAGGTGTCTCTGACGGAGGCCGAGATGAAGACGGCGCTG   720
L  E  S  I  P  G  F  K  V  S  L  T  E  A  E  M  K  T  A  L     152
GAGGAGGTCGGCTGCTGTATCGTAGGCCAGACCGCCGACATCGTACCTGCTGACAGACGC   780
E  E  V  G  C  C  I  V  G  Q  T  A  D  I  V  P  A  D  R  R     172
ATGTACGCCGCAAGAGACGTCGCTTCAACCGTCAAATCTGTGCCGCTCATCGTCTCGTCC   840
M  Y  A  A  R  D  V  A  S  T  V  K  S  V  P  L  I  V  S  S     192
ATCATCAGCAAGAAGGCTGCGGAAACCGTGAGCGGGCTGGTGCTCGACGTCAAGTTCGGC   900
I  I  S  K  K  A  A  E  T  V  S  G  L  V  L  D  V  K  F  G     212
GGAGGAGCCTTCATGAAGACCCAGGAGGAGGCAGGGGCGCTGGCCAAGAAAATGGTGGAT   960
G  G  A  F  M  K  T  Q  E  E  A  G  A  L  A  K  K  M  V  D     232
GTGGCCAACGGCGTGGGCATGGCCACGACGGCCCTCCTGACCACGATGGATATCCCGCTC   1020
V  A  N  G  V  G  M  A  T  T  A  L  L  T  T  M  D  I  P  L     252
GGCAGGGCCATCGGCAACGCCCTCGAGGTGCGGGAGTCGCTGGAGTGTCTTCGGGGCAAC   1080
G  R  A  I  G  N  A  L  E  V  R  E  S  L  E  C  L  R  G  N     272
GGACCGGAGGACCTTGAGGAGCTCGTAACGCACCTGGGCGGAGAATTACTGCTGGGTGCG   1140
G  P  E  D  L  E  E  L  V  T  H  L  G  G  E  L  L  L  G  A     292
GGAGCGGCCTCTACGCTGGATGAAGCTCGCCAGAAGCTGGCCAAGGCTCTGAGGGATGGC   1200
G  A  A  S  T  L  D  E  A  R  Q  K  L  A  K  A  L  R  D  G     312
AGTGCCAGAACGGCTTTCTGCAATATGATACAGAAGCAGGGTGTCACCAAGAGTGTAGCA   1260
S  A  R  T  A  F  C  N  M  I  Q  K  Q  G  V  T  K  S  V  A     332
GAGGCACTGTGCGGCAATGTTCCCGACTACTCCCATCTACCTTCCTCGGCTCATGTCACT   1320
E  A  L  C  G  N  V  P  D  Y  S  H  L  P  S  S  A  H  V  T     352
GCCGTCAAAGCTGCTTCCTCAGGAGTGCTAGTTGGTATGGATGCCATGACTATGGCGAAG   1380
A  V  K  A  A  S  S  G  V  L  V  G  M  D  A  M  T  M  A  K     372
ATCAGTTTAGAACTCGGGGCTGGCAGGAACAAGGTCGGCGACCCGATCAACTACAGCGTG   1440
I  S  L  E  L  G  A  G  R  N  K  V  G  D  P  I  N  Y  S  V     392
GGCATAATGCTCATTAAGGTGGTCGGCGAGAGCGTGAAGGAAGGCGAGACGTGGGCAGAG   1500
G  I  M  L  I  K  V  V  G  E  S  V  K  E  G  E  T  W  A  E     412
CTGCACCACGATTCCTCACTGCCACCCACCCTCCTACAGAGGATGCAGGGAGCCGTCACC   1560
L  H  H  D  S  S  L  P  P  T  L  L  Q  R  M  Q  G  A  V  T     432
ATCAAGGCGTCGGCGGAAGCATGCAAGCCCTCGCGCGTTGCCGCTCGCGTTGTCTAGTG   1620
I  K  A  S  A  E  A  C  K  P  S  R  V  A  A  R  V  V  *        450
CTCGAAACCCTCAGTTGCAGAGGCCAAGATCCAGCTCGCACTGTCAGTCTCCTAACTTGA   1680
```

GATCTAGTATCCATGCAAAATACTTTATGTTGCACAAATTATGTTAGGGCCTGTTCGCA*T* 1740
*GAGCGGACATAATCGGC*GGGCGCGCGTACGCAAACGTAATAGTCCGGTGGGCTAGTGCTC 1800
CTAACAGCAGTTCCCTTTGTAACGTCACCTTCACAGGTGCCTGACGTTGCGGACAGGTCT 1860
TTAACCGTTTGAACCTTCTACTTTTTTATTTTCTAATTATATGGGAAACTGGAGACGGCT 1920
CTCTTCCGGGTTCATCGTCATACTAGTCACTGACCCCAAATATACTAGTACATGGGGAAG 1980
GGGACTGTTATTTTCACGTGGCACAATCTTCACATCGCCAGTTCTTTCTCCCAAAACCAT 2040
ATTCATCGTCAAATAGAAAGTATACCTATAGATACTTATATTCGGTGCATATATGCCAGT 2100
TGCTCTTCCCAGTATATAACTGTAACCATCATTAACTATTCCAGGTACCTGCTGGCGCTC 2160
ATCAGGATTTTTTCTTTACGAAACTAACTTAATTGCGTTGTTGTTGTTTTTTTTGGTTCA 2220
CAAATATGATTTAATGTGTTAGTTTTTTGGTTCACAAGAATTATTTGATTTTGTTATGTA 2280
AATCAGCCGGTGTGTTAATCGGTCTGCAGTTAATACGATTGTTCTGATTCTGGCTTCCAG 2340
TCACTGTGGTTCTCTTTATGATTTCAGCTTACTACTTCATAACAATCTTCTACATAGTGA 2400
TTTCTGATATTTTTGATGTTTGCACAAGCTATTTAATTGGATGGTATGAAATTACACATG 2460
ACCTGCATATTAGTCATCTATTGGCTGACCTAACCGATTCATTCCACTCCTTGAAAGGGA 2500
AATAACTCATTCTGTACAACTCTTTACTATGCAACTGGCCAACGTCATCCTCACATTCCA 2560
GTCTATGTAGATTTAACAAGAAATATTTTTGTGATTGAAATATTTTTTGTTGAGTTTTAC 2620
TATTGTGTTGTAATGGGAAATTAATAGTTTTTGTATGATATTGAAATTTACACAATGTCT 2680
CTCCTTGTTCATGAA**AATAAA**ACTATTGCAGTGCTATAAAAAAAAAAAAAAAAAAAAAA 2739

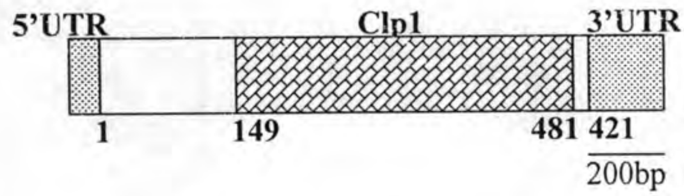
**Figure 3.58** The full length cDNA sequences of *PMECGF I.* Start and stop codons are illustrated in boldface and underlined. The 5′ECGF, 3′ECGF and internal primers are underline and italicized. The polyA additional signal is illustrated in boldface.

### 10. *ATP/GTP binding protein* (*PMATP/GTP*)

The full length cDNA of *PMATP/GTP* was obtained from sequencing of both 5′ and 3′ of the EST significantly matched to a homologue of *ATP/GTP* binding. The full length cDNA of *PMATP/GTP* was 1547 bp in length. The ORF of *PMATP/GTP* was 1266 bp encoding a polypeptide of 421 amino acids. The 5′ and 3′UTRs of *ATP/GTP* were 68 and 213 bp (excluding the poly A tail; Figure 3.59 and 3.60).

The closest sequence to the full length *PMATP/GTP* was *cleavage/polyadenylation factor in subunit clp1* of *Aedes aegypti* (E-value = 2e-144). The calculated pI was 5.98 with the molecular weight of 46517.8 dalton and the signal peptide was not found in this putative protein.

The pre-mRNA cleavage complex II protein, Clp1 domain was found at positions 149 - 481 (Figure 3.59). This protein family consists of several pre-mRNA cleavage complex II Clp1 (or HeaB) proteins. Six different protein factors are required *in vitro* for the 3' end formation of mammalian pre-mRNAs by endonucleolytic cleavage and polyadenylation. Clp1 is a subunit of cleavage complex IIA, which is required for cleavage, but not for polyadenylation of pre-mRNA




| Domain | Position | E-value |
|---|---|---|
| Clp1 | 149-418 | 5.50e-157 |

**Figure 3.59** Diagram illustrating the full length cDNA of *ATP/GTP* of *P. monodon* The pre-mRNA cleavage complex II protein Clp1 domain was found in this transcript. A scale bar is 200 bp in length.

CACAGCTATCGCCTCTTTCCTTCGGTTTATCATTGAAAATCCTAGAAAAGGTATACAGTG 60
GATGCAAA**ATG**TCTGAAAAGGAGTTCCGTTTGGACCCAGACAGCGAACTCAGGTTTGAGG 120
M S E K E F R L D P D S E L R F E V 18
TTGAGGGCAAGCAGGATACAGTTGATCTCACATTGGTAAATGGCAAAGCTGAAGTGTTTG 180
E G K Q D T V D L T L V N G K A E V F G 38
GCACAGAACTAGCACCAGATAAGCCGTATACCTTCTTCCCTGGAGCAAAAGTAGCAGTGT 240
T E L A P D K P Y T F F P G A K V A V F 58
TCACCTGGCATGGCTGTGTTCTCAAGCTGTCTGGCCCAACAGAGGGAACATATGTGGCAA 300
T W H G C V L K L S G P T E G T Y V A K 78
AGGAAACCCCAATGGTTATGTACCTTAATACACATGCATGCTTGGAGCGGTTGCGCAGGC 360
E T P M V M Y L N T H A C L E R L R R H 98
ATGCAGATGAAGGTCTTAGCCGAGGAGAGGATACCAGAGGACCAGTAACCATGGTTGTTG 420
A D E G L S R G E D T R G P V T M V V G 118
GTCCAGGAGATGTAGGAAAGTCAACTCTGTGCAGAATTCTCCTCAAC**TATGCTGTGCGGA** 480
P G D V G K S T L C R I L L N Y A V R M 138
**TGGGACG**TCGGCCCATCTTTGTAGATTTGGATATTGGACAAGGATCCATTGCAATTCCAG 540
G R R P I F V D L D I G Q G S I A I P G 158
GAACTATAGGTGCACTGCTTGTTGAGAGAGCTGCTGATGTTGGGGAAGGCTTTTCTCAGG 600
T I G A L L V E R A A D V G E G F S Q E 178
AGGCTCCTCTTGTGTACAACTTTGGGCACCTGAGTCCTGCATCCAACATGACTCTGTACA 660
A P L V Y N F G H L S P A S N M T L Y N 198
ACATCCTTGTCTCACGCATGGCTGCCACCATTCAAGATAAGATGGTTGGCAATAGAAAAG 720
I L V S R M A A T I Q D K M V G N R K V 218
TTGCTGCCTCAGGGGTGGTTATCAACACATGTGGCTGGATTAAAAATGAAGGTTACAAGA 780
A A S G V V I N T C G W I K N E G Y K S 238
GTCTGACACATGTTGCGCAGGCCTTTGAGGTTGACGTGATCATTGTCCTTGACCAGGAGA 840
L T H V A Q A F E V D V I I V L D Q E R 258
GACTATATAATGAACTCGTCAGAGATATTCCCTTCATTAGGGTTGTTTTCTTGCCCAAGA 900
L Y N E L V R D I P F I R V V F L P K S 278
GTGGAGGTGTGGTAGAGAGAACTCAATCCATGAGGGCATCTGCTCGTGATGACCGCATCA 960
G G V V E R T Q S M R A S A R D D R I R 298
GAGAATACTACTATGGACTGCGCACCAAGTACCACCCACACAGTTTTGAGGTGAAGATGA 1020
E Y Y Y G L R T K Y H P H S F E V K M S 318
GCACGCTCCAGATATACAAGATTGGTGCTCCAGCCCTCCCAGATTCCTGCATGCCTGCTG 1080
T L Q I Y K I G A P A L P D S C M P A D 338
ATATGAAGGTGGATGATCACATGACCAAACTTGTGCCTGTAGAGCCTAGTGTAAAGTTAA 1140
M K V D D H M T K L V P V E P S V K L K 358

```
AGCATCACATGTTGGCTGTAAGTCTTGCAACAGAACCCGAGGACCTCTTGACCTCCAATG     1200
 H  H  M  L  A  V  S  L  A  T  E  P  E  D  L  L  T  S  N  V      378
TCGCTGGATTTATTTGCGTACTTGACGTTGATGAAGACCTCAAAGTCATGAAGGTCTTGT     1260
 A  G  F  I  C  V  L  D  V  D  E  D  L  K  V  M  K  V  L  S      398
CTCCACAACCAAAGCCACTCCCAAAAACAATTTTGATTTTGACGGAAATTCAGTTCATGG     1320
 P  Q  P  K  P  L  P  K  T  I  L  I  L  T  E  I  Q  F  M  D      418
ACTCATCGTAAAATGAAGAGTGGTTTTGTAATTTCTCGTAAATCATGTGGTATTACATTT     1380
 S  S  *                                                         420
TTTTTTTTCTCTCTCTCTCTCTTCTCCTTTTTTCCTATATTTTCTTTGATATAGCAGTTA     1440
TTTATATTGTGTAATTTTTGTATACAGATGATTTCTGCTGATTTTTGGTGTAACTTAAAA     1500
TTTTTTATTAAATAAGTATTTTACAAGAAAAAAAAAAAAAAAAAAAAA                  1547
```

**Figure 3.60** The full length cDNA sequences of *ATP/GTP* of *P. monodon.* Start and stop codons are illustrated in boldface and underlined. The internal sequencing primer is boldfaced.

### 11. *Ovarian lipoprotein receptor* (*PMOVLP*)

Several fragments were obtained from 3′RACE-PCR of a *PMOVLP* primer (Figure 3.61). The largest product of approximately 3000 bp in length was cloned and sequenced for both directions. Primer walking was applied for sequencing of the entire 3′RACE-PCR fragment using 2 internal primers. Nucleotide sequences of the original EST and 3′RACE-PCR were assembled and analyzed.

The complete ORF of *PMOVLP* was still not obtained. The combined nucleotide of *PMOVLP* was 3437 bp in length. The partial ORF of *PMOVLP* was 3318 bp translated to 1105 amino acids. The 3′UTRs of *PMOVLP* was 116 bp (excluding the poly A tail). (Figure 3.62 and 3.63).

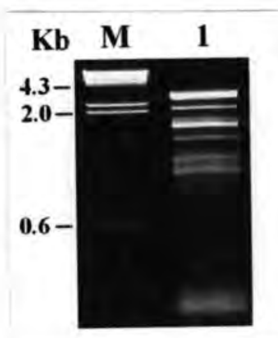
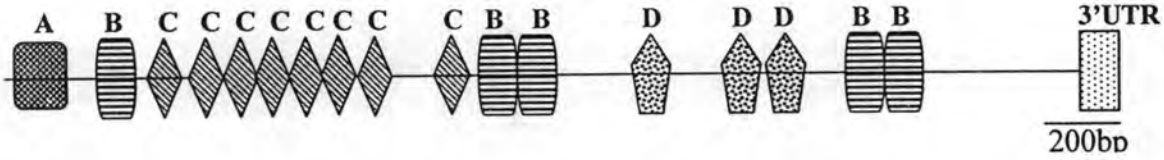



**Figure 3.61** The primiary 3′RACE-PCR product of *PMOVLP* (lane 1). The λ–*Hin*d III (lane M) was used as the markers.

The closest sequence to *PMOVLP* was *ovarian lipoprotein receptor* of the green tiger shrimp, *Penaeus semisulcatus* (E-value = 0.00) followed by the *vitellogenin receptor* of *Blattella germanicaat* (E-value = 5e-130).

Several functionally important domains were found such as, the internal repeat 3 domain, epidermal growth factor domain, low-density lipoprotein receptor domain class A and LY domain were found in this partial transcript.




| Domain | Begin | End | E-value |
|---|---|---|---|
| internal repeat 3 | 7 | 61 | 7.71e-05 |
| EGF_like | 103 | 141 | 5.62e+00 |
| LDLa | 144 | 182 | 8.74e-10 |
| LDLa | 185 | 224 | 9.81e-13 |
| LDLa | 226 | 263 | 4.27e-13 |
| LDLa | 264 | 301 | 4.05e-14 |
| LDLa | 302 | 341 | 1.44e-10 |
| LDLa | 347 | 389 | 9.98e-05 |
| LDLa | 390 | 429 | 3.31e-10 |
| LDLa | 452 | 490 | 1.68e-11 |
| EGF_like | 491 | 527 | 4.56e+00 |
| EGF_like | 528 | 567 | 2.15e-03 |
| LY | 635 | 676 | 4.17e+01 |
| LY | 747 | 789 | 3.79e-06 |
| LY | 790 | 830 | 9.69e+00 |
| EGF_like | 857 | 895 | 2.62e+00 |
| EGF_like | 892 | 933 | 2.00e-01 |

**Figure 3.62** Diagram illustrating of partial *PMOVLP* with 3'UTR and internal repeat 3 domain (A), epidermal growth factor domain (B), Low-density lipoprotein receptor domain class A (C) and LY domain.(D) (200bp: 1 centimeter)

LDLa domains are cysteine-rich repeats that play a central role in mammalian cholesterol metabolism. The *N*-terminal type A repeats in LDL receptor bind the lipoproteins.

Low density lipoprotein (LDL) is the major cholesterol-carrying lipoprotein of plasma. The LDL-receptor class A domain contains 6 disulphide-bound cysteines and a highly conserved cluster of negatively charged amino acids. LY domain is the type "B" repeats in low-density lipoprotein (LDL) receptor that plays a central role in mammalian cholesterol metabolism.

Epidermal growth factors (EGF)-like domain shares a repeat pattern involving a number of conserved cys residues. Growth factors are involved in cell recognition and division found frequently in nature, particularly in extracellular proteins.

```
AGCGTGGTCGCGGCCGAGGTACTGGTGTGACCACGTCCTGGGACGTATCGAGAGCGTGGC   60
 A  W  S  R  P  R  Y  W  C  D  H  V  L  G  R  I  E  S  V  A    20
CGAAGACGGGACGGACCGACGGGTATTCATGGAAAACGTGAAGAGCCCCGTCTCGGTCCT   120
 E  D  G  T  D  R  R  V  F  M  E  N  V  K  S  P  V  S  V  L    40
GGTGACTCGTTCGCACATCATGTGGTCTGAGGAACGGACTTCTCTCATCTACTCCGCTTC   180
 V  T  R  S  H  I  M  W  S  E  E  R  T  S  L  I  Y  S  A  S    60
CAAATTAGATAACAGTTCTGTGCGGATGATGTCTCTTGATCTGGGCGTTCCGGAAAATGG   240
 K  L  D  N  S  S  V  R  M  M  S  L  D  L  G  V  P  E  N  G    80
AGAACGTTCCCTGAAAATTCTGGAAGTCGGGTGGAAAATTCCCGAGCAACTTGCAGCGAC   300
 E  R  S  L  K  I  L  E  V  G  W  K  I  P  E  Q  L  A  A  T    100
GAACCACCCGTGTCTGCAGAGTAACGAGAGCTGTAGTCAGTTGTGCCTTGGAGACGACTT   360
 N  H  P  C  L  Q  S  N  E  S  C  S  Q  L  C  L  G  D  D  F    120
CAACGAAAAGGTGTGCGCTTGCAGTTTCGGCTACAAACTCCAGGTGGACAGAAGAACGTG   420
 N  E  K  V  C  A  C  S  F  G  Y  K  L  Q  V  D  R  R  T  C    140
TGAATCCGTCCAGTGTAACGACATCCAGTTTCACTGTTTTCGATCGCACACTTGTATTCC   480
 E  S  V  Q  C  N  D  I  Q  F  H  C  F  R  S  H  T  C  I  P    160
TCGTTCATGGAAATGTGACTTGACCCCAGACTGCAAAGACGGCGAAGATGAAGAGGACTG   540
 R  S  W  K  C  D  L  T  P  D  C  K  D  G  E  D  E  E  D  C    180
CAAGGAGTCAACGACGTGTAAGGAGAAAGAGTTCCGGTGCTCTACGGGTTCATGCATTAA   600
 K  E  S  T  T  C  K  E  K  E  F  R  C  S  T  G  S  C  I  N    200
CAAGCTGTGGACGTGTGACGGTGTGCACGATTGCGAGGATGGCTCTGACGAGAGACTCGA   660
 K  L  W  T  C  D  G  V  H  D  C  E  D  G  S  D  E  R  L  D    220
TGAATGCACGAATGTGACATGTAGCAGCGTACATTGGAAGTGCAAGTCAGGCATGTGCAT   720
 E  C  T  N  V  T  C  S  S  V  H  W  K  C  K  S  G  M  C  I    240
TCCGAAGATGTGGGTTTGTGACCAAGAGAAAGAGTGCGACGATGGGTCGGACGAGACTGA   780
 P  K  M  W  V  C  D  Q  E  K  E  C  D  D  G  S  D  E  T  E    260
GTGCGTTACTTCCTGCCCTGACCATAAAGTCGCTTGTAGAGATGGAAAATGTGTACCAAA   840
 C  V  T  S  C  P  D  H  K  V  A  C  R  D  G  K  C  V  P  K    280
GGTGTGGAAATGTGACGGCGACAAGGACTGCCTGGACGGAAGTGACGAGGAAAATTGTCC   900
 V  W  K  C  D  G  D  K  D  C  L  D  G  S  D  E  E  N  C  P    300
GGTGGAGTGTGCGAACAATGAGTTCACCTGCAGCAACAAAAACTGTGTACCCCACGATGC   960
 V  E  C  A  N  N  E  F  T  C  S  N  K  N  C  V  P  H  D  A    320
CAAGTGCGATGGCGAGGATGACTGTGGCGACGGCTCTGATGAGGCACTTCCTTGGTGCCA   1020
 K  C  D  G  E  D  D  C  G  D  G  S  D  E  A  L  P  W  C  Q    340
GCCTCCGGACCCTCCTGTAACGTGCCCCAGCGGCCAGATTTTGTGTGAACGCCACGACGT   1080
 P  P  D  P  P  V  T  C  P  S  G  Q  I  L  C  E  R  H  D  V    360
TTCTTCCCCGCGTGTCTGTATACAACTGAATAACGTATGCAACGGTGTCCGGGACTGCCC   1140
```

S S P R V C I Q L N N V C N G V R D C P 380
CTTAGGCGAAGACGAGGACTGTGATAACTGCGCCCGCCACGAGTTCAGCTGCTTATCACG 1200
L G E D E D C D N C A R H E F S C L S R 400
TGGCTGCATC***CCAAGAGGGTGGATGTGTGAT***GGGAAGAGGACTGCACTGACGGCTCCGA 1260
G C I P R G W M C D G E E D C T D G S D 420
CGAGAGCCAGGCAGCCGGTTGCATGATTGCACAAGGCAATGACACCGTTGATCTAAGCCT 1320
E S Q A A G C M I A Q G N D T V D L S L 440
GAATGGTAGCGATGGTAAGGCAGCTCCAGTGCCAGTCTGTGGAATACACGAGTTCGAGTG 1380
N G S D G K A A P V P V C G I H E F E C 460
TGGGATTGGCGGCTGCATAGCGTCGCGTCTTGTGTGTGATGGCTCGGCGGACTGTCTTGA 1440
G I G G C I A S R L V C D G S A D C L D 480
TGGCTCTGATGAAGGCAGCTTGTGCGCAAAAAGTTGTCTGGGTAATGGCGGGTGCCAACA 1500
G S D E G S L C A K S C L G N G G C Q H 500
CACATGCAAAGAAGGTCCCAAAAATCGCATTTGCTCCTGTTGGAAGGGATTCCAACTCGC 1560
T C K E G P K N R I C S C W K G F Q L A 520
CGAGGATCAGATTAGCTGCATTGATGTGAAGGAATGCGACGATGAGGCCACCTGCAGCCA 1620
E D Q I S C I D V K E C D D E A T C S Q 540
AAAGTGCGAAGAAAGACATGGCTACCACTTGTGCTCCTGCCTACCCGGGTATACTCTTAG 1680
K C E E R H G Y H L C S C L P G Y T L R 560
ACCTGACAGACGCTCTTGCAAACCAGCAGGTGGCGACGAATATGTGGTCTTGGTGCATC 1740
P D R R S C K P A G G D E Y V V L V H P 580
TGGGTCCATCCTGAATATGTCCCGCACCTTCCATCTTGCTGACAAAGTGGCGATGCCCCC 1800
G S I L N M S R T F H L A D K V A M P P 600
TCATGTTCAGTTTTCGTCTGTTGAGTTTACGCCCGAGTCCCATAATTTCGTTTATGCTGA 1860
H V Q F S S V E F T P E S H N F V Y A D 620
CAAAGCCCATGGAGTTATCGGGAAGATGAGCATGGACGGCGTAGTGACAATACTCTTTA**A** 1920
K A H G V I G K M S M D G V V T I L F K 640
**GCACAGAAAGCGTCCTCAG**GGTCTCTCCTTGGACCCCATTAGCAACAGCGTTTATTTCTC 1980
H R K R P Q G L S L D P I S N S V Y F S 680
CGAACAGTTCAGTAAAGCTGAAGTTGTGGATAATGGCTTGATAAGAGTGCGCAGGGAGCC 2040
E Q F S K A E V V D N G L I R V R R E P 700
GAGTGCTGCTGGGACTTATTCTGTGATAATGGTTTGTGGATGGAGGGCGACAAGGATTG 2100
S A A G T Y S V I M V C G M E G D K D C 720
CAGCATGGTGTACCAATCACATGGTGGAGAGATCCCGGCAATCCGTGTTGCCCCAATGGC 2160
S M V Y Q S H G G E I P A I R V A P M A 740
AAGACGACTCTTCTTCTGCGCTAACAACGTGGCGCAGGACGAAGCAAAAATTTTCACCTC 2220
R R L F F C A N N V A Q D E A K I F T S 760
GGATATGGATGGCACATCGGCTCGAATTCTCAGCCATAAGGTTGTGAAGTGTGGTGACCT 2280
D M D G T S A R I L S H K V V K C G D L 780
GGCAGTGGATGAGGCAAAGGAGCGAGTCTACTGGACGGATCTCTCCCGTAACGTTATCGA 2340
A V D E A K E R V Y W T D L S R N V I E 800
GTCCGTCAAATGGTCAGGCGAAGGCCATCGTGTTGTACAAGAAAATGTACACACGCCAAT 2400
S V K W S G E G H R V V Q E N V H T P I 820
TGGACTAGCCTTGATTGAAGACTGGGTGCTGTGGCTGGACACGCACCAGCACCAAATAAT 2460
G L A L I E D W V L W L D T H Q H Q I I 840
CAAGTGTAACAAGTACAAGATGGGTATGTGTGACCACCACACCATGGGCACTGCCGGCTT 2520
K C N K Y K M G M C D H H T M G T A G L 860
AGCTTTGACTGTTCAGCATCGTTTAAGAATGGAGAGTCCATTGATTGGAGACTGCAGAGT 2580
A L T V Q H R L R M E S P L I G D C R V 900
AAAAAAAAACTGCACTCACCACTGCATGATTCAAATGGGCAAAAAGGCCAGCTGTATGTG 2640
K K N C T H H C M I Q M G K K A S C M C 920
CAAAGTTGGCTACATCTCTGCACCCAGCCGTCCTAACGAGTGTATCAGGATGAAATCCTG 2700
K V G Y I S A P S R P N E C I R M K S C 940
CGACCACAGCCCGTGTCAAGGCAAAGGTATATGCGAGTCGCACTCCGAATCAGAGTTCAT 2760
D H S P C Q G K G I C E S H S E S E F I 960
TTGCAGGTGTCCTGAAGGCCGTGAAGGGTCCCTGTGCGAGGTGGCCAAGACGCCCACAGC 2820
C R C P E G R E G S L C E V A K T P T A 980
AGACAACAGCGGCAGCGGCAGCAGCGCAACCTTAGGCGTGTGCCTCTTCCTCCTCTTTTT 2880
D N S G S G S S A T L G V C L F L L F F 1000
CGGTGCCCTCCTCTTTGGGCTTTATTGGTATCGAAAGCAACCTTTCCCCTTTTGGAAGGG 2940
G A L L F G L Y W Y R K Q P F P F W K G 1020
AAAAGGAGGGCAACTTCGCAAGAGATGCTTCAAAGCCAATCAGACCCTACGTTTCGCCAA 3000
K G G Q L R K R C F K A N Q T L R F A N 1040
CCCAGGTTTTGGCATCATTTCCCCCACCACTGTGCCCAACGGAAATGGGACGTCCAGCAC 3060
P G F G I I S P T T V P N G N G T S S T 1060
CAACAGCAACACCATCCCCTCAACCCCGCCTGTCTTGGGAGGTTCTCACAACTTCGAAAA 3120
N S N T I P S T P P V L G G S H N F E N 1080
CCCTTTCTTTAAAACTGATGAGCACGTGCCGGACACGAGTGCGGACTCGGCCATAGTGAG 3180

```
 P  F  F  K  T  D  E  H  V  P  D  T  S  A  D  S  A  I  V  S     1100
CACAGCCGACTCGACCTCCATCAACATCGCTCCCCATCAGGGGGATCTGACCCCGCCACA   3240
 T  A  D  S  T  S  I  N  I  A  P  H  Q  G  D  L  T  P  P  Q     1120
GAACGTACTGAAGCCACCGGTAGAGAAGAGGGTCGAGTGGGATCTCTCTCCTTTCCAGCC   3300
 N  V  L  K  P  P  V  E  K  R  V  E  W  D  L  S  P  F  Q  P     1140
TTTGCAGCCTCAGGTTGAGGCAGTGCATTGGAATAGATCTATTGATTTAGATGCACTTCT   3360
 L  Q  P  Q  V  *                                              1160
TATAATTGTTATGTTTAATGTTATAGCCTAATATATCTTTCTAAACTAAAAAAAAAAAAA   3420
AAAAAAAAAAAAAAAAA                                              3437
```

**Figure 3.63** The partial cDNA sequences of *PMOVLP*. The stop codon are illustrated in boldface and underlined. The 3'RACE-PCR primer is underlined, boldfaced and italicized whereas the internal primers are boldfaced

### 3.4 Semiquantitative RT-PCR of *female sterile, adipose differentiation related protein, nuclear autoantigenic sperm protein, ovarian lipoprotein receptor, 3-oxoacid CoA transferase, dolichyl diphophooligosaccharide protein glycotransferase* and *aspartase aminotransferase* upon induction by 5-HT treatment

Total RNA extracted from ovaries of normal and 5-HT-treated juvenile (approximately 4-month-old) *P. monodon* females were determined by spectrophotometry and by agarose gel electrophoresis (Figure 3.64). The ratio of $OD_{260}/ OD_{280}$ of the extracted RNA was 1.8 – 2.0 indicating its acceptable purity for further used for the reverse transcription. Agarose gel electrophoresis showed sharp ribosomal RNA bands reflecting good quality of total RNA isolated from ovaries of juvenile *P. monodon*. The first strand cDNA was successfully synthesized as revealed by 1.2% agarose gel electrophoresis (Figure 3.65).

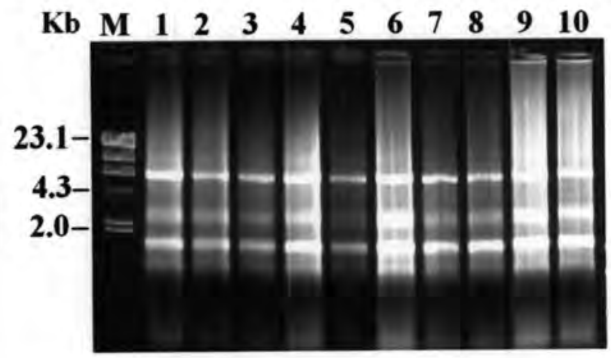



**Figure 3.64** A 1.0% ethidium bromide-stained agarose gel showing the quality of total RNA extracted from ovaries of juvenile *P. monodon*. Lane M = λ-*Hin*d III, Lanes 1-10 = total RNA individually extracted from ovaries of each *P. monodon*.


**Figure 3.65** A 1.0% ethidium bromide-stained agarose gel showing the quality of the first strand cDNA synthesized from ovaries of juvenile *P. monodon*. Lane M = λ-*Hin*d III. Lane 1-10 = the first strand cDNA from ovaries of each *P. monodon*.

### 3.4.1 Optimization of semi-quantitative RT-PCR conditions

The first strand cDNA of ovaries of juvenile shrimp single injected with the normal saline (A0) and 5-HT for 12, 24, 48 and 72 hours (A12, A24, A48 and A72, respectively; $N = 5$ for each group) and those of juvenile shrimp repeatedly injected with normal saline (B0) and 5-HT for 12, 24, 48 and 72 hours (B12, B24, B48 and B72, respectively; $N = 5$ for each group) after the first injection were used as template for a time-course analysis of homologues of *female sterile (PMFS), adipose differentiation related protein (PMADRP), nuclear autoantigenic sperm protein (PMNASP), ovarian lipoprotein receptor (PMOVLP), 3-oxoacid CoA transferase, aspartate amino transferase (PMAST)* and *dolichyl diphosphooligosaccharide protein glycotransferaase* (*PMDDPG*) using semiquantitative RT-PCR analysis. This technique requires optimization of several parameters including concentration of primers, $MgCl_2$, and the number of PCR cycles.

Primers for the target genes were designed. *EF-1*$\alpha$ was used as the control. The preliminary RT-PCR was carried out using the standard conditions and the annealing temperature of 53°C as previously used during screening of gene expression patterns of various genes.

#### 3.4.1.1 Optimization of the primer concentration

RT-PCR of each gene was carried out with fixed components except primer concentrations (0.1, 0.15, 0.20, 0.25 and 0.30 μM). Lower concentrations may result in non-quantitative amplification whereas higher concentrations of primer may leave a large amount of unused primers which could give rise to non-specific amplification products. The suitable concentration of primers for each gene is shown by Table 3.8.

#### 3.4.1.2 Optimization of the $MgCl_2$ concentration

The optimal concentration of $MgCl_2$ (between 1.0, 1.5, 2.0, 2.5 and 3.0 mM) for each primer pair was carefully examined using the amplification conditions with the optimaized primer concentration. The concentration of $MgCl_2$ that gave the highest yields and specificity for each PCR product was chosen (Table 3.8).

### 3.4.7.3 Optimization of the cycle numbers

The number of amplification cycles was important because the product reflecting the expression level should be measured quantitatively before reaching a plateau amplification phase. At the plateau stage, transcripts initially present at different levels may give equal intensity of the amplification products.

In this experiment, RT-PCR of each gene was performed using the conditions that primers and $MgCl_2$ concentrations were optimized for 18, 20, 23 and 25 cycles. The number of cycles that gave the highest yield before the product reached a plateau phase of amplification was chosen (Table 3.8).

**Table 3.8** Optimal primer and $MgCl_2$ concentrations and the number of PCR cycles for semiquantitative analysis of genes in *P. monodon*

| Transcript | Expected amplicons (bp) | Primer concentration (μM) | $MgCl_2$ concentration (mM) | PCR cycles |
|---|---|---|---|---|
| *Female sterile* | 296 | 0.15 | 1.5 | 20 |
| *Adipost differential related protein* | 267 | 0.20 | 1.0 | 23 |
| *Ovarian lipoprotein receptor* | 354 | 0.15 | 2.0 | 25 |
| *Nuclear autoantigenic sperm protein* | 301 | 0.20 | 2.0 | 25 |
| *Aspartate aminotransferase* | 334 | 0.20 | 2.0 | 25 |
| *3-Oxoacid CoA transferase* | 303 | 0.20 | 2.0 | 25 |
| *Dolichyl diphosphooligo-saccharide protein glycotransferaase* | 233 | 0.20 | 2.0 | 25 |
| *Elongationfactor 1-α* | 500 | 0.125 | 1.5 | 23 |

### 3.4.2 Semi-quantitative RT-PCR analysis

#### 3.4.2.1 *PMFS*

The expression level of *PMFS* in juvenile *P. monodon* upon single injection with 5-HT (group A) was greater than that of double injection (group B, $P$ <0.05) and the expression level of both treatment was significantly higher than the control ($P$ <0.05).

Within the single injection group, the expression level of *PMFS* was initially up-regulated at 12 hour post treatment (12 hpt; 0.965 ± 0.029, $P$ < 0.05) and further increased at 24 hpt (1.323 ± 0.045, $P$ < 0.05). The highest expression of *PMFS* was observed at 48 hpt (1.470 ± 0.070, $P$ < 0.05). The expression was slightly reduced but still significant from that of the control A0 at 72 hpt (1.348 ± 0.119, $P$ < 0.05) (Figure 3.66 and 3.67; Table 3.9).

Within the double injection group, the expression of *PMFS* was significantly up-regulated at 12 hpt (0.803 ± 0.015, $P$ <0.05) and increased at 24 and 48 hpt (0.930 ± 0.057 and 0.904 ± 0.043, respectively; $P$ < 0.05). The expression of *PMFS* was reduced but still significant to that of the control Bo at 72 hpt (0.853 ± 0.015, $P$ < 0.05) (Figures 3.66 and 3.67; Table 3.9).

#### 3.4.2.2 *PMADRP*

The expression level of *PMADRT* in juvenile *P. monodon* upon single and double injection with 5-HT was comparable ($P$ > 0.05). This gene was only up-regulated at 48 hpt (0.661 ± 0.057, $P$ < 0.05) and returned to the normal levels at 72 hpt (0.550 ± 0.019, $P$ > 0.05). Repeat injection of 5-HT did not affect the expression of *PMADRT* when compared with the control B0 (0.426 ± 0.057, $P$ > 0.05) (Figures 3.68 and 3.69; Table 3.9).

#### 3.4.2.3 *PMNASP*

The expression level of *PMNASP* in juvenile *P. monodon* upon single injection with 5-HT (A) was lower than that of double injection (B, $P$ < 0.05) and the expression level of both treatment was significantly higher than the control ($P$ < 0.05).

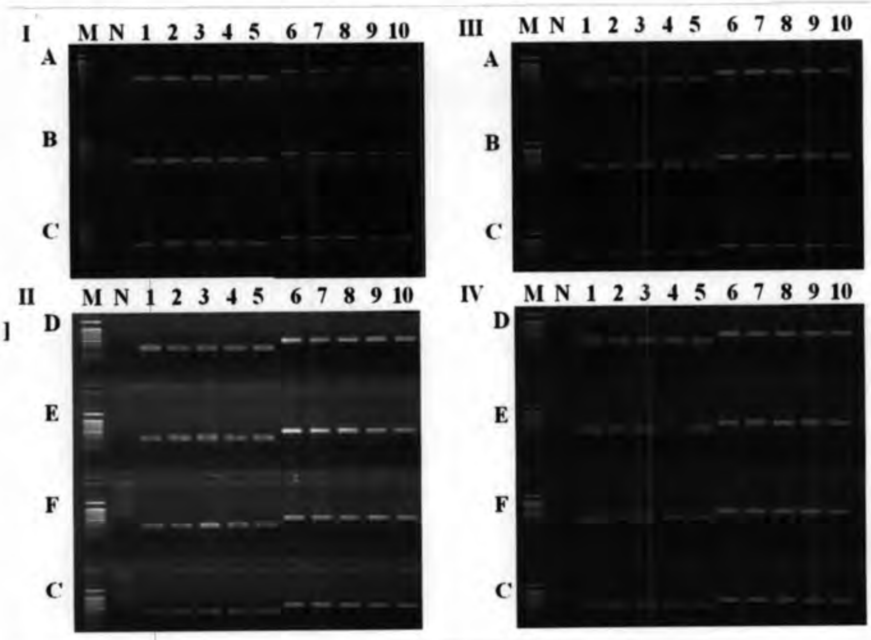



**Figure 3.66** A 1.6% ethidium bromide-stained agarose gel showing the expression level of *PMFS* (lanes 1 - 5) and *EF-1α* (lanes 6 - 10) of single (I and II) and double (III and IV) injection of 5-HT for 12, 24, 48 and 72 hours post injection (B, D, E, F). The positive control (0.85% NaCl) for the single injection (A, C and G of I and II; control A) and the double injection (A, C and G of III and IV; control B) was also included. Lane M = 100 bp DNA ladder.

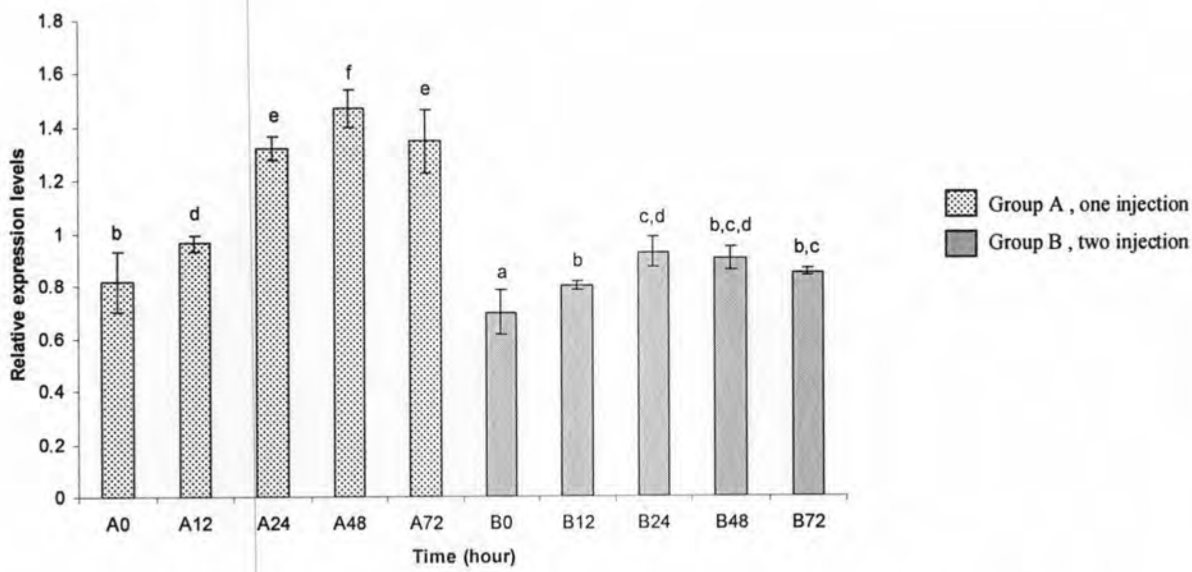



**Figure 3.67** Histograms showing the time-course relative expression levels of *PMFS* for 12, 24, 48 and 72 hours post treatment of 5-HT. The positive control (0.85% NaCl) for the single (A0) and double (B0) injection was also analyzed. The same letters indicate that the expression levels were not significantly different ($P > 0.05$).

Within the single injection group, the expression of *PMNASP* was only up-regulated at 12 hpt (0.799 ± 0.033, $P < 0.05$) and returned to the normal level since 24 hpt ($P > 0.05$) (Figures 3.70 and 3.71; Table 3.9).

Within the double injection group, the expression of *PMNASP* was initially up-regulated to the highest level at 24 hpt (0.799 ± 0.033, $P < 0.05$) and still significantly different from the control B0 at 48 and 72 hpt (0.738 ± 0.062 and 0.831 ± 0.049, respectively, $P < 0.05$) (Figure 3.70 and 3.71; Table 3.9).

### 3.4.2.4 *PMOVLP*

The expression level of *PMOVLP* in juvenile *P. monodon* upon single injection with 5-HT (A) was significantly greater than that of double injection (B, $P < 0.05$) and the expression level of both treatment was significantly higher than the control ($P < 0.05$).

Within the former group, the gene expression of *PMOVLP* was initially up-regulated at 24 hpt (0.773 ± 0.034, $P < 0.05$). The expression level was constant at 48 and 72 hpt (0.816 ± 0.053 and 0.752 ± 0.047, respectively) and significant from that of the control A0 (0.580 ± 0.071, $P < 0.05$) (Figure 3.72 and 3.73; Table 3.9).

Within the latter group, the expression of *PMOVLP* was up-regulated to the highest level within the treatment at 24 hpt (0.634 ± 0.049, $P < 0.05$). The expression level of *PMOVLP* was slightly lowered at 48 and 72 hpt (0.576 ± 0.060 and 0.538 ± 0.088, respectively) but still significant to that of the control B0 (0.435 ± 0.077, $P < 0.05$) (Figures 3.72 and 3.73; Table 3.9).

### 3.4.2.5 *3-oxoacid CoA transferase*

The expression level of *3-oxoacid CoA transferase* in juvenile *P. monodon* upon single injection with 5-HT (A) was approximately equal to that of double injection (group B, $P > 0.05$) and the expression level of both treatment was significantly higher than that of the control ($P < 0.05$).

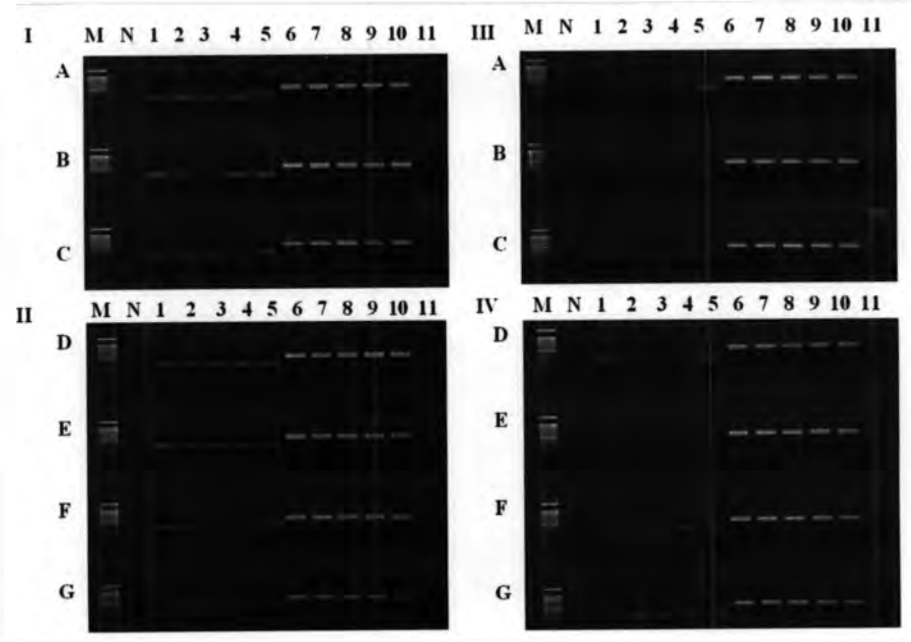



**Figure 3.68** A 1.6% ethidium bromide-stained agarose gel showing the expression level of *PMADRP* (lanes 1 - 5) and *EF-1α* (lanes 6 - 10) of single (I and II) and double (III and IV) injection of 5-HT for 12, 24, 48 and 72 hours post injection (B, D, E, F). The positive control (0.85% NaCl) for the single injection (A, C and G of I and II; control A) and the double injection (A, C and G of III and IV; control B) was also included. Lane M = 100 bp DNA ladder.

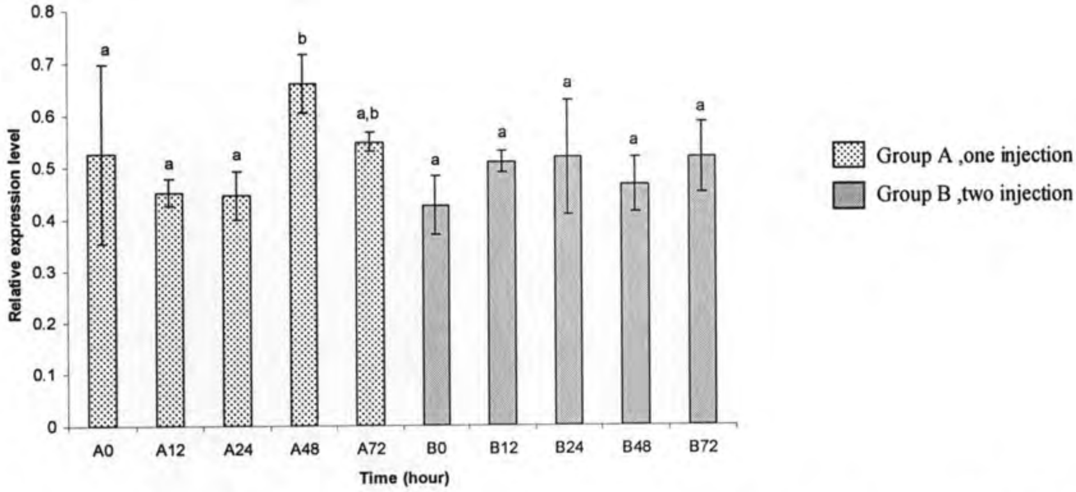



**Figure 3.69** Histograms showing the time-course relative expression levels of *PMADRP* for 12, 24, 48 and 72 hours post treatment of 5-HT. The positive control (0.85% NaCl) for the single (A0) and double (B0) injection was also analyzed. The same letters indicate that the expression levels were not significantly different ($P > 0.05$).

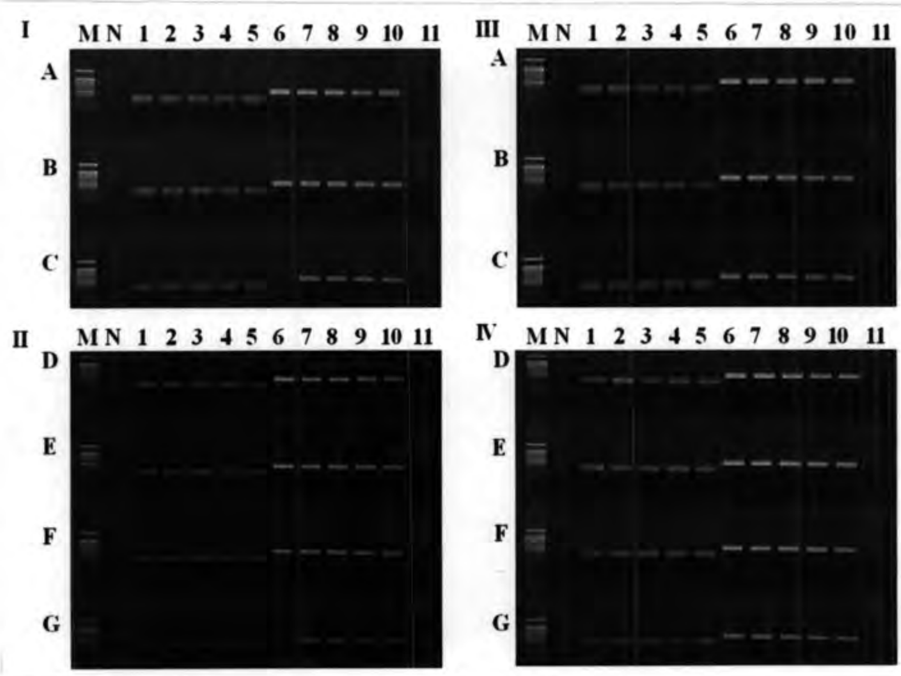



**Figure 3.70** A 1.6% ethidium bromide-stained agarose gel showing the expression level of *PMNASP* (lanes 1 - 5) and *EF-1α* (lanes 6 - 10) of single (I and II) and double (III and IV) injection of 5-HT for 12, 24, 48 and 72 hours post injection (B, D, E, F). The positive control (0.85% NaCl) for the single injection (A, C and G of I and II; control A) and the double injection (A, C and G of III and IV; control B) was also included. Lane M = 100 bp DNA.ladder

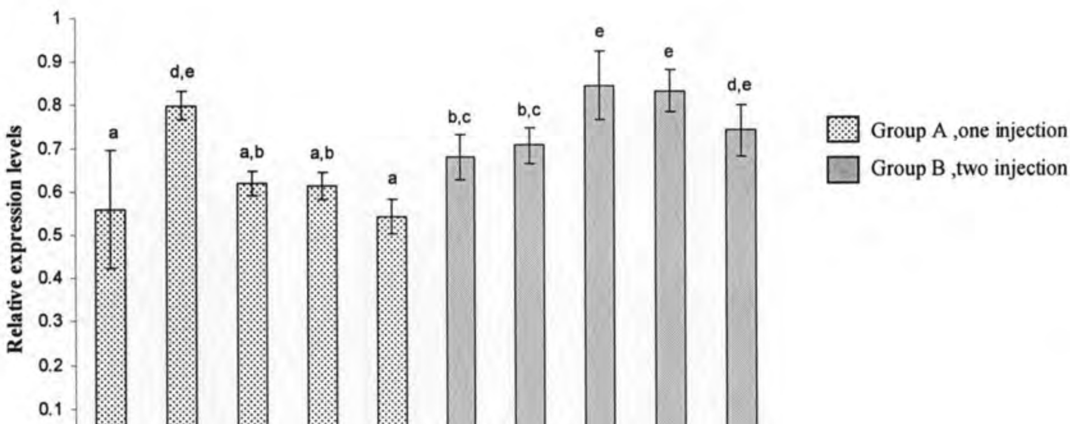



**Figure 3.71** Histograms showing the time-course relative expression levels of *PMNASP* for 12, 24, 48 and 72 hours post treatment of 5-HT. The positive control (0.85% NaCl) for the single (A0) and double (B0) injection was also analyzed. The same letters indicate that the expression levels were not significantly different ($P > 0.05$).

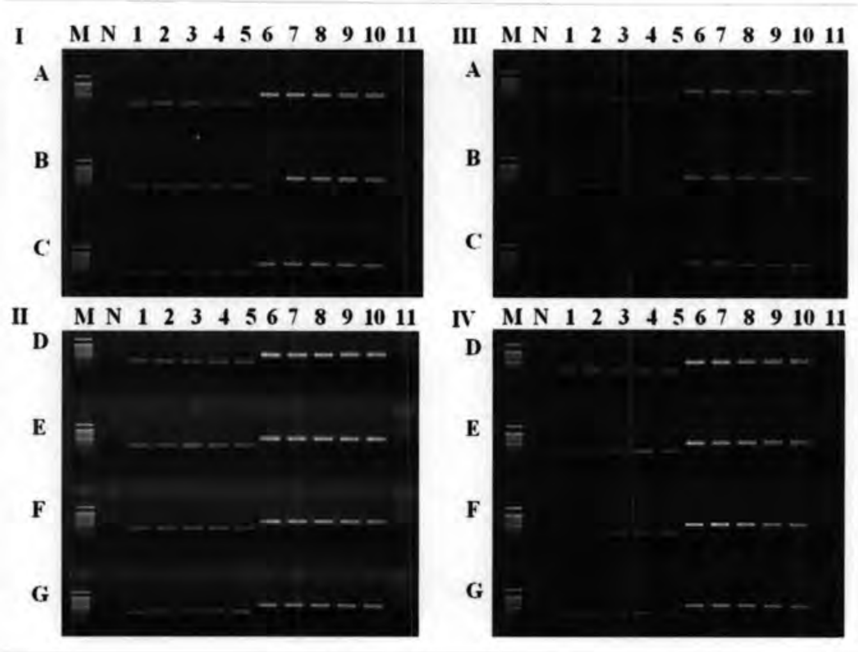



**Figure 3.72** A 1.6% ethidium bromide-stained agarose gel showing the expression level of *PMOVLP* (lanes 1 - 5) and *EF-1α* (lanes 6 - 10) of single (I and II) and double (III and IV) injection of 5-HT for 12, 24, 48 and 72 hours post injection (B, D, E, F). The positive control (0.85% NaCl) for the single injection (A, C and G of I and II; control A) and the double injection (A, C and G of III and IV; control B) was also included. Lane M = 100 bp DNA ladder.

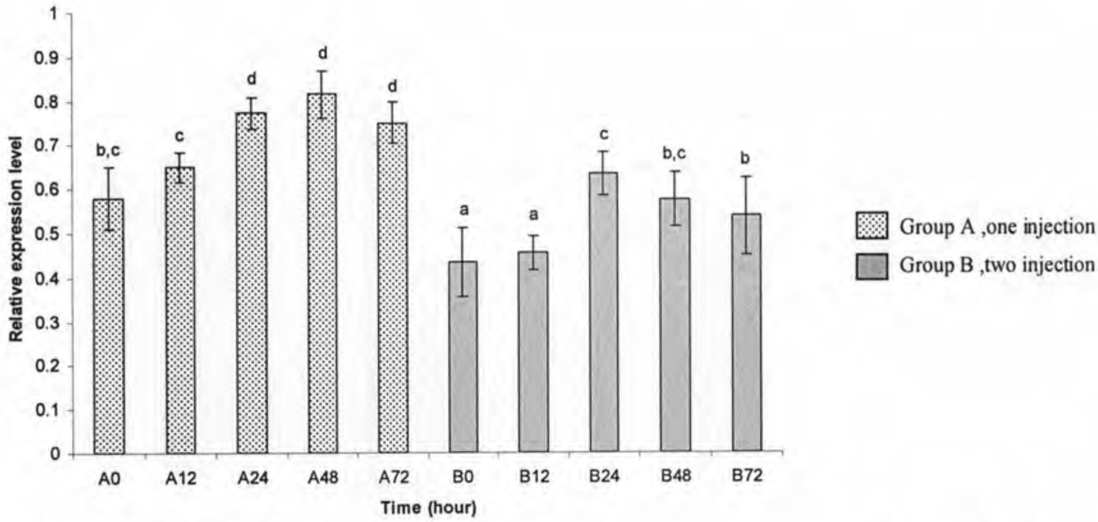



**Figure 3.73** Histograms showing the time-course relative expression levels of *PMOVLP* for 12, 24, 48 and 72 hours post treatment of 5-HT. The positive control (0.85% NaCl) for the single (A0) and double (B0) injection was also analyzed. The same letters indicate that the expression levels were not significantly different ($P > 0.05$).

Within the single injection group, the expression of *3-oxoacid CoA transferase* was initially up-regulated at 12 hpt (0.389 ± 0.014, $P$ <0.05) and further increased at 24 and 48 hpt (0.516 ± 0.043 and 0.450 ± 0.045, respectively, $P$ <0.05). The expression level was reduced but still significant to that of the control A0 at 72 hpt (0.410 ± 0.026 compared to 0.309 ± 0.042, $P$ <0.05). (Figure 3.74 and 3.75; Table 3.9).

Likewise, 5-HT provided similar effects on levels of expression of *3-oxoacid CoA transferase* in the double injection group. The gene expression was initially up-regulated at 12 hpt (0.405 ± 0.027, $P$ <0.05) and further increased at 24 and 48 hpt (0.536 ± 0.054 and 0.481 ± 0.057, respectively, $P$ <0.05). The expression level was still significant to that of the control B0 at 72 hpt (0.482 ± 0.091 compared to 0.328 ± 0.017, $P$ <0.05) (Figure 3.74 and 3.75; Table 3.9).

#### 3.4.2.6 *PMDDPG*

The expression level of *PMDDPG* in juvenile *P. monodon* upon single (A) and double injection (B) with 5-HT was not different ($P$ > 0.05) and the expression level of both treatment was significantly higher than that of the control ($P$ <0.05).

Within the single injection group, the expression of *PMDDPG* was initially up-regulated at 12 hpt (0.665 ± 0.084, $P$ <0.05) and further increased to the highest level at 24 hpt (0.702 ± 0.086, $P$ <0.05). The expression level was slightly reduced but still significant from that of the control A0 at 48 hpt (0.661 ± 0.048, $P$ <0.05) and returned to the normal level at 72 hpt (0.538 ± 0.028, $P$ >0.05) (Figure 3.76 and 3.77; Table 3.9).

Similarly, the expression of *PMDDPG* within the double injection group was initially up-regulated at 12 hpt (0.691 ± 0.037, $P$ <0.05) and further increased to the highest level at 24 hpt (0.759 ± 0.125, $P$ <0.05). The expression was reduced to the same level as that of the control B0 at 48 hpt (0.539 ± 0.102 compared to 0.440 ± 0.127, $P$ >0.05) and returned to the normal level at 72 hpt (0.497 ± 0.049, $P$ >0.05) (Figure 3.76 and 3.77; Table 3.9).

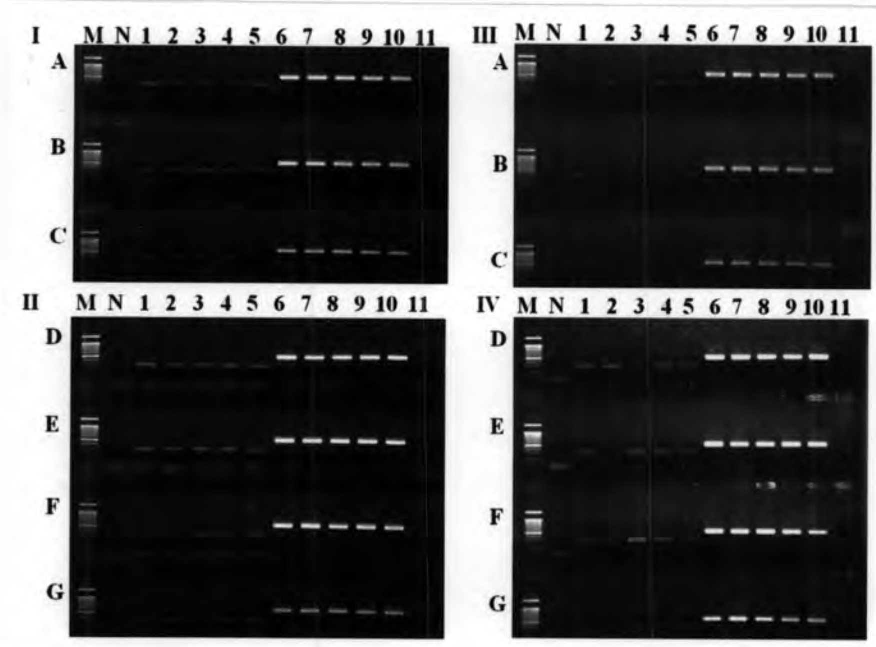



**Figure 3.74**. A 1.6% ethidium bromide-stained agarose gel showing the expression level of *3-oxoacid CoA transferase* (lanes 1 - 5) and *EF-1α* (lanes 6 - 10) of single (I and II) and double (III and IV) injection of 5-HT for 12, 24, 48 and 72 hours post injection (B, D, E, F). The positive control (0.85% NaCl) for the single injection (A, C and G of I and II; control A) and the double injection (A, C and G of III and IV; control B) was also included. Lane M = 100 bp DNA ladder.

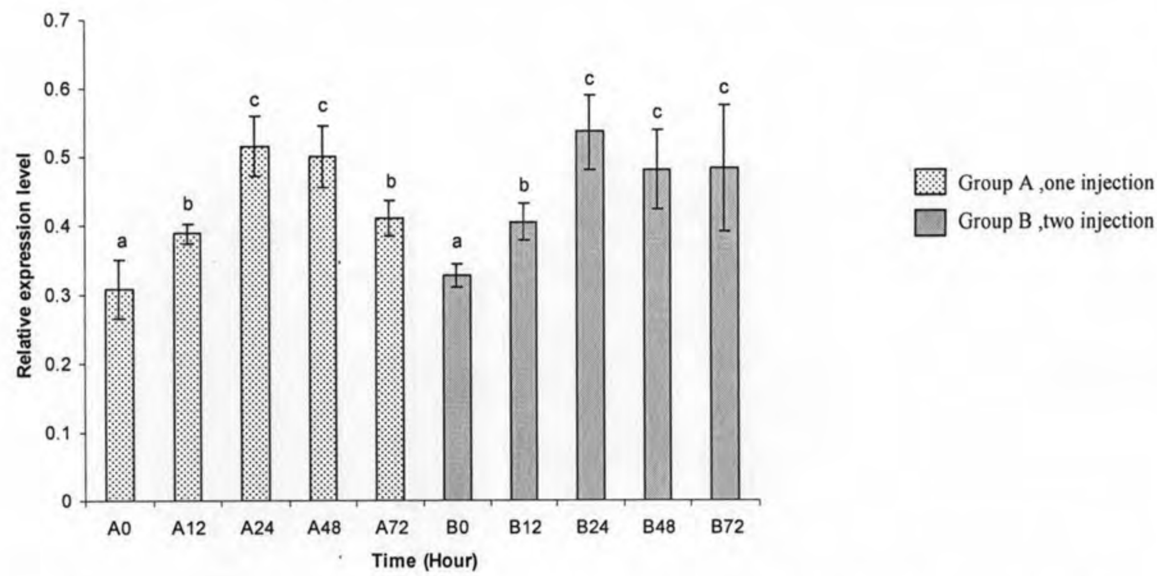



**Figure 3.75** Histograms showing the time-course relative expression levels of *3-oxoacid CoA transferase* for 12, 24, 48 and 72 hours post treatment of 5-HT. The positive control (0.85% NaCl) for the single (A0) and double (B0) injection was also analyzed. The same letters indicate that the expression levels were not significantly different ($P > 0.05$).

### 3.4.2.7 *PMAST*

The expression level of *PMAST* in juvenile *P. monodon* upon single (A) and double injection (B) with 5-HT was similar ($P > 0.05$) and the expression level of both treatment was significantly higher than that of the control ($P < 0.05$).

Within the single injection group, the expression level of *PMAST* was initially up-regulated at 12 hpt (0.480 ± 0.067, $P < 0.05$) and further increased to the highest level at 24 hpt (0.505 ± 0.046, $P < 0.05$). The expression was slightly reduced but still significant from that of the control A0 at 48 and 72 hpt (0.482 ± 0.043 and 0.482 ± 0.030, $P < 0.05$) (Figure 3.78 and 3.79; Table 3.9).

The expression level of *PMAST* within the double injection group was initially up-regulated at 12 hpt (0.471 ± 0.033, $P < 0.05$). The highest expression level was observed at 48 hpt (0.501 ± 0.071, $P < 0.05$). The expression was reduced but still significantly different from that of the control B0 at 72 hpt (0.481 ± 0.055 compared to 0.384 ± 0.046, $P < 0.05$) (Figure 3.78 and 3.79; Table 3.9).

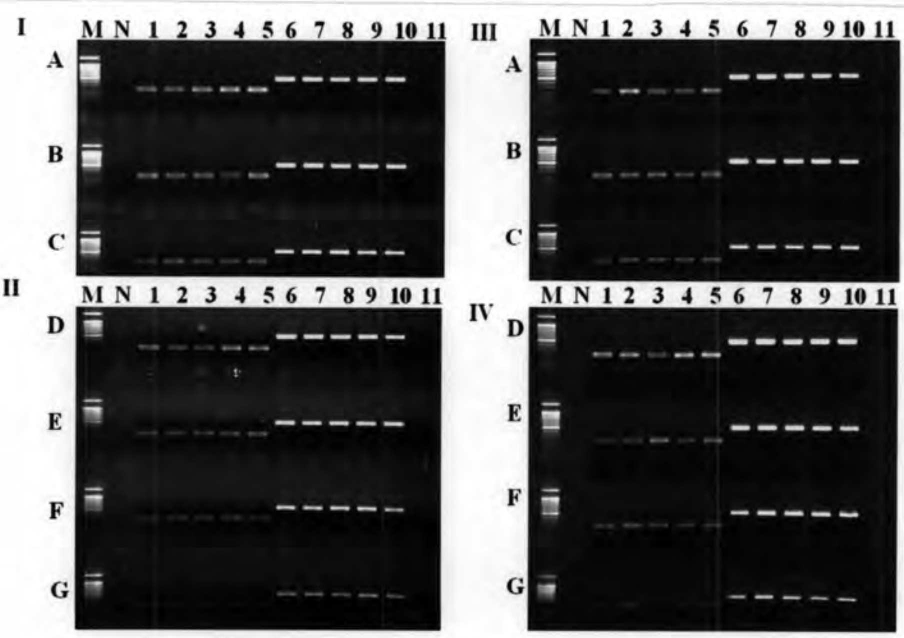



**Figure 3.76** A 1.6% ethidium bromide-stained agarose gel showing the expression level of *PMDDPG* (lanes 1 - 5) and *EF-1α* (lanes 6 - 10) of single (I and II) and double (III and IV) injection of 5-HT for 12, 24, 48 and 72 hours post injection (B, D, E, F). The positive control (0.85% NaCl) for the single injection (A, C and G of I and II; control A) and the double injection (A, C and G of III and IV; control B) was also included. Lane M = 100 bp DNA ladder.

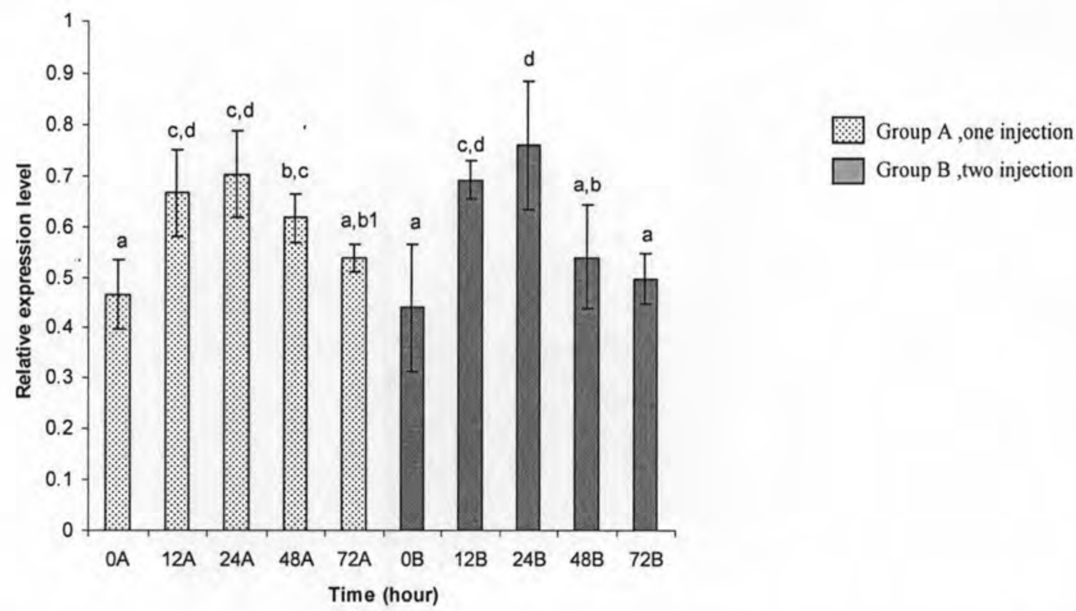



**Figure 3.77** Histograms showing the time-course relative expression levels of *PMDDPG* for 12, 24, 48 and 72 hours post treatment of 5-HT. The positive control (0.85% NaCl) for the single (A0) and double (B0) injection was also analyzed. The same letters indicate that the expression levels were not significantly different ($P > 0.05$).

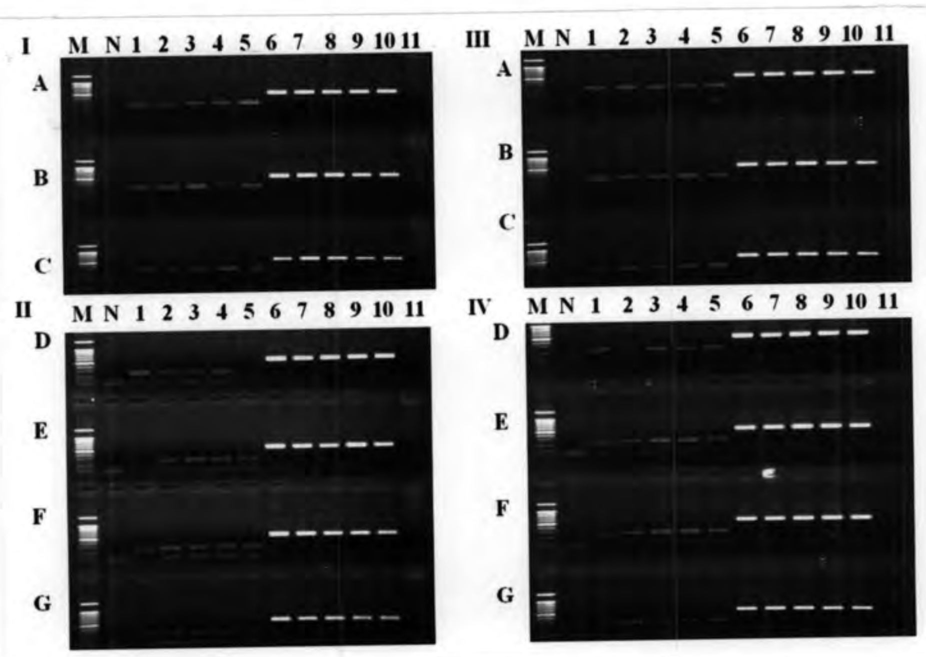



**Figure 3.78** A 1.6% ethidium bromide-stained agarose gel showing the expression level of *PMAST* (lanes 1 - 5) and *EF-1α* (lanes 6 - 10) of single (I and II) and double (III and IV) injection of 5-HT for 12, 24, 48 and 72 hours post injection (B, D, E, F). The positive control (0.85% NaCl) for the single injection (A, C and G of I and II; control A) and the double injection (A, C and G of III and IV; control B) was also included. Lane M = 100 bp DNA ladder.

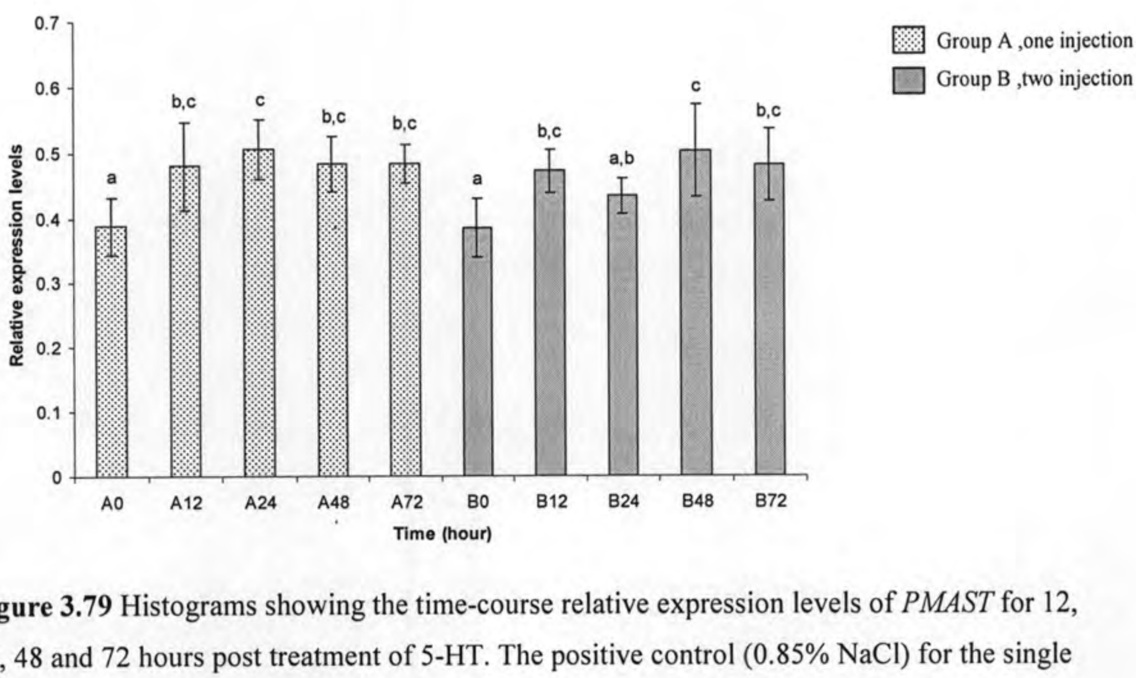



**Figure 3.79** Histograms showing the time-course relative expression levels of *PMAST* for 12, 24, 48 and 72 hours post treatment of 5-HT. The positive control (0.85% NaCl) for the single (A0) and double (B0) injection was also analyzed. The same letters indicate that the expression levels were not significantly different ($P > 0.05$).

**Table 3.9** A time-course analysis of expression levels of various genes using semiquantitative RT-PCR. The same superscripts between different time interval data are not significantly different ($P > 0.05$)

| Gene | Mean relative expression level | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Control A** | **12A** | **24A** | **48A** | **72A** | **Control B** | **12B** | **24B** | **48B** | **72B** |
| *PMFS* | 0.818±0.113$^{b}$ | 0.965±0.029$^{e}$ | 1.323±0.045$^{c,d}$ | 1.470±0.070$^{f}$ | 1.348±0.119$^{e}$ | 0.698±0.084$^{a}$ | 0.803±0.015$^{b}$ | 0.930±0. 057$^{c,d}$ | 0.904±0.043$^{b,c,d}$ | 0.853±0.015$^{b,c}$ |
| *PMADRP* | 0.526±0.171$^{a}$ | 0.452±0.025$^{a}$ | 0.447±0.047$^{a}$ | 0.661±0.057$^{b}$ | 0.550±0.019$^{a,b}$ | 0.426±0.057$^{a}$ | 0.511±0.021$^{a}$ | 0.521±0.110$^{a}$ | 0.467±0.053$^{a}$ | 0.519±0.068$^{a}$ |
| *PMNASP* | 0.559±0.135$^{a}$ | 0.799±0.033$^{d,e}$ | 0.619±0.028$^{a,b}$ | 0.612±0.032$^{a,b}$ | 0.541±0.042$^{a}$ | 0.676±0.053$^{b,c}$ | 0.704±0.040$^{b,c}$ | 0.842±0.079$^{e}$ | 0.831±0.049$^{e}$ | 0.738±0.062$^{c,d}$ |
| *PMOVLP* | 0.580±0.071$^{b,c}$ | 0.649±0.033$^{c}$ | 0.773±0.034$^{d}$ | 0.816±0.053$^{d}$ | 0.752±0.047$^{d}$ | 0.435±0.077$^{a}$ | 0.456±0.039$^{a}$ | 0.634±0.049$^{c}$ | 0.576±0.060$^{b,c}$ | 0.538±0.088$^{b}$ |
| *PM 3-Oxoacid CoA transferase* | 0.309±0.042$^{a}$ | 0.389±0.014$^{c}$ | 0.516±0.043$^{c}$ | 0.450±0.045$^{b}$ | 0.410±0.026$^{b}$ | 0.328±0.017$^{a}$ | 0.405±0.027$^{b}$ | 0.536±0.054$^{c}$ | 0.481±0.057$^{c}$ | 0.482±0.091$^{c}$ |
| *PMDDPG* | 0.467±0.068$^{a}$ | 0.665±0.084$^{c,d}$ | 0.702±0.086$^{c,d}$ | 0.616±0.048$^{b,c}$ | 0.538±0.028$^{a,b}$ | 0.440±0.127$^{a}$ | 0.691±0.037$^{c,d}$ | 0.759±0.125$^{d}$ | 0.539±0.102$^{a,b}$ | 0.497±0.049$^{a}$ |
| ***PMAST*** | 0.387±0.044$^{a}$ | 0.480±0.067$^{b,c}$ | 0.505±0.046$^{c}$ | 0.482±0.043$^{b,c}$ | 0.482±0.030$^{b,c}$ | 0.384±0.046$^{a}$ | 0.471±0.033$^{b,c}$ | 0.433±0.027$^{a,b}$ | 0.501±0.071$^{c}$ | 0.481±0.055$^{b,c}$ |

*The expression of *EF-1α* was normalized to 1.00. The relative expression level of the target genes was determined as the signal ratio between the target gene and *EF-1α*.