CHAPTER IV

RESULTS

4.1 DNA Extraction

High molecular weight genomic DNA was located at approximately 48.5 kb, indicating good quality of extracted genomic DNA (Fig.4.1).

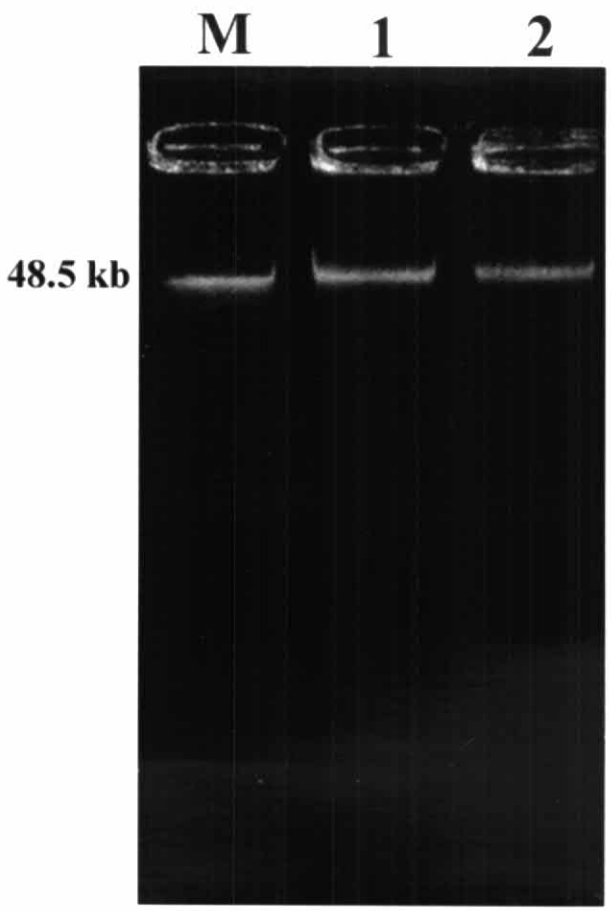


Figure 4.1 A 1.2% ethidium bromide stained agarose gel showing the quality of genomic DNA extracted from the muscle of *L. subviridis* (lane 1 and 2). Lane M is 100 ng of undigested lambda DNA marker.

4.2 RNA Extraction

The extracted RNA showed 28s and 18s rRNA indicating good quality of RNA (Fig. 4.2). RNA was used for mRNA purification for RACE cDNA synthesis and as the template for 1st strand cDNA synthesis, which was then used as the template for amplifying partial and full-length cDNA sequences as well as semiquantitative RT-PCR analysis.

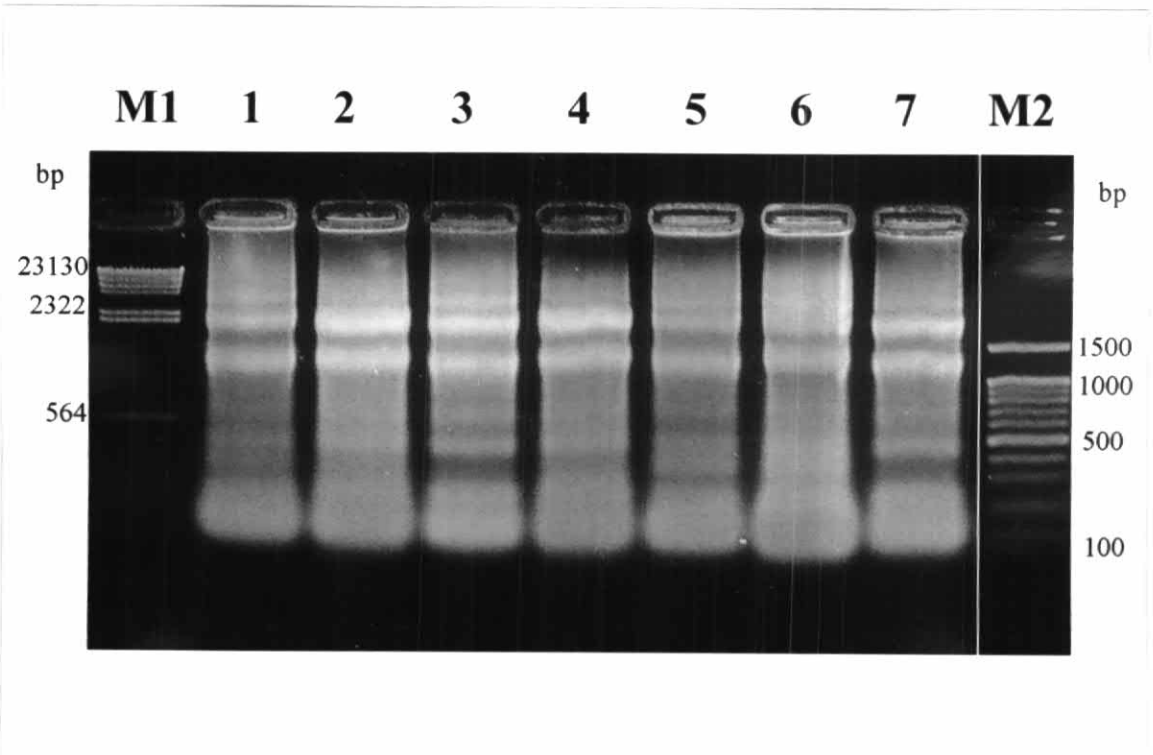


Figure 4.2 A 1.2 % ethidium bromide stained agarose gel showing the quality of RNA extracted from the liver of *L. subviridis*. Lane 1 – 7 is total liver RNA of normal, 0, 0.05, 0.1, 0.5, 1 and 5 mg/kg E_2 treated (6 days) mullet, respectively. Lane M1 and M2 are λ*Hin*dIII and 100 bp DNA marker, respectively.

4.3 Cloning and characterization of estrogen receptor (ER) genes

4.3.1 PCR amplification of estrogen receptor (ER) gene

Partial coding sequence of ER gene was amplified from 1st strand cDNA template of the liver. DNA fragment at the size of 100 bp was obtained (Fig. 4.3). After sequencing (see in appendix B (*ER* FR/59) and BLAST analysis (see in appendix C), this PCR product was homologous with estrogen receptor alpha of flathead mullet (*Mugil cephalus*) with E-value of 6e-36.

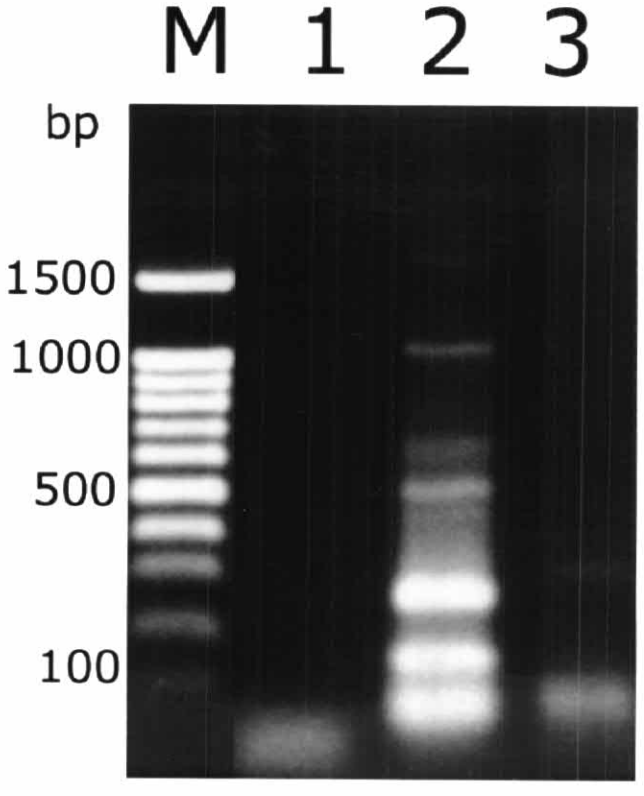


Figure 4.3 PCR products of ER gene separated in 1.2% agarose gel. PCR products amplified from liver. Lane M is 100 bp markers, lane 1 is negative control, lane 2 is positive control, and lane 3 is PCR product of ER gene.

4.3.2 Amplification of 5' and 3' cDNA ends of ERα gene by RACE-PCR

DNA fragment at the size of 1.76 kb was obtained from 3'RACE PCR amplification in liver (Fig. 4.4).

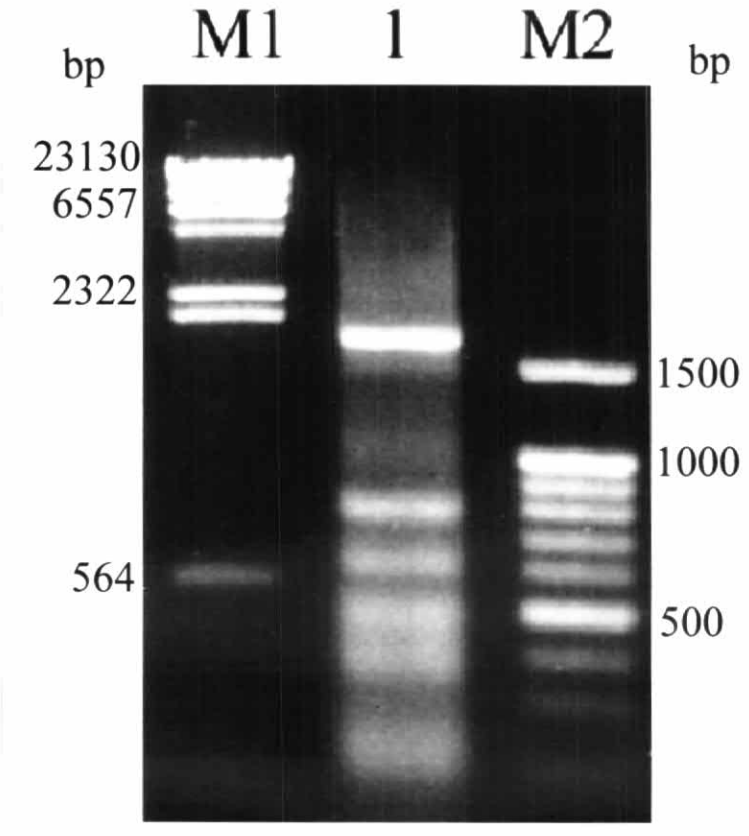


Figure 4.4 3' RACE PCR products of ER gene separated in 1.2% agarose gel. PCR products amplified from 3'RACE cDNA of liver (lane M1 is λ*Hin*dIII, lane M2 is 100 bp markers and lane 1 is 3' RACE PCR product of *ER*).

PCR product with a few different sizes was obtained. Band at the size of 1.76 kb which was close to an expected size was cloned. Several white colonies were checked for the insert size by colony PCR. Some examples of colony PCR products were shown in Fig. 4.5 A and B. Clone No.36 (insert size of 1.76 kb), 32 and 5 (insert size lower than 1.76 kb) were cultured for plasmid extraction and sequencing (see in appendix B; 3'*ER*/5, 32 and 36). After BLAST analysis (see in appendix C), it was found that 3'*ER*/5, 32 and 36 was homologous with estrogen receptor alpha of European sea bas (*Dicentrarchus labrax*), flathead mullet (*Mugil cephalus*) and thicklip grey mullet (*Chelon labrosus*) with E-value as 7e-101, 0.0 and 0.0, respectively. Because of *ER* FR/59 sequence used for designing *ER* F-RACE primer was DNA binding domain which was very highly conserved region between *ER*α and *ER*β. 3' RACE PCR product of ER gene could amplify 2 isoforms. Therefore, PCR product from 3'*ER* colonies with the insert size of 1.76 kb were digested with *Hin*dIII (Fig. 4.6).

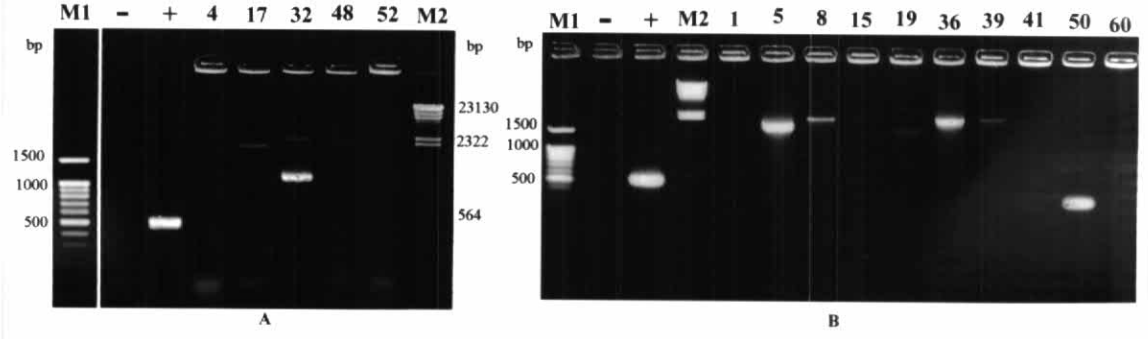


Figure 4.5 A and B; Colony PCR product of 3'RACE *ER* separated in 1.2% agarose gel. Lane M1 is 100 bp DNA marker, lane M2 is λ*Hin*dIII, - is negative control, + is positive control and other lanes is 3'RACE *ER* colony PCR product; number of lane represent number of clone.

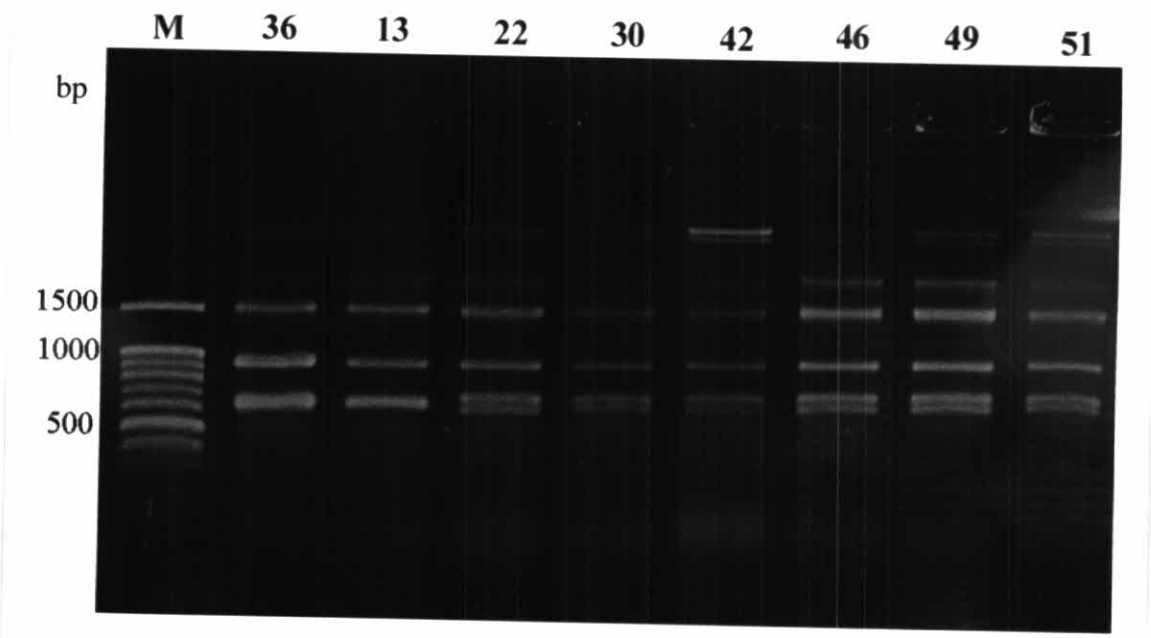
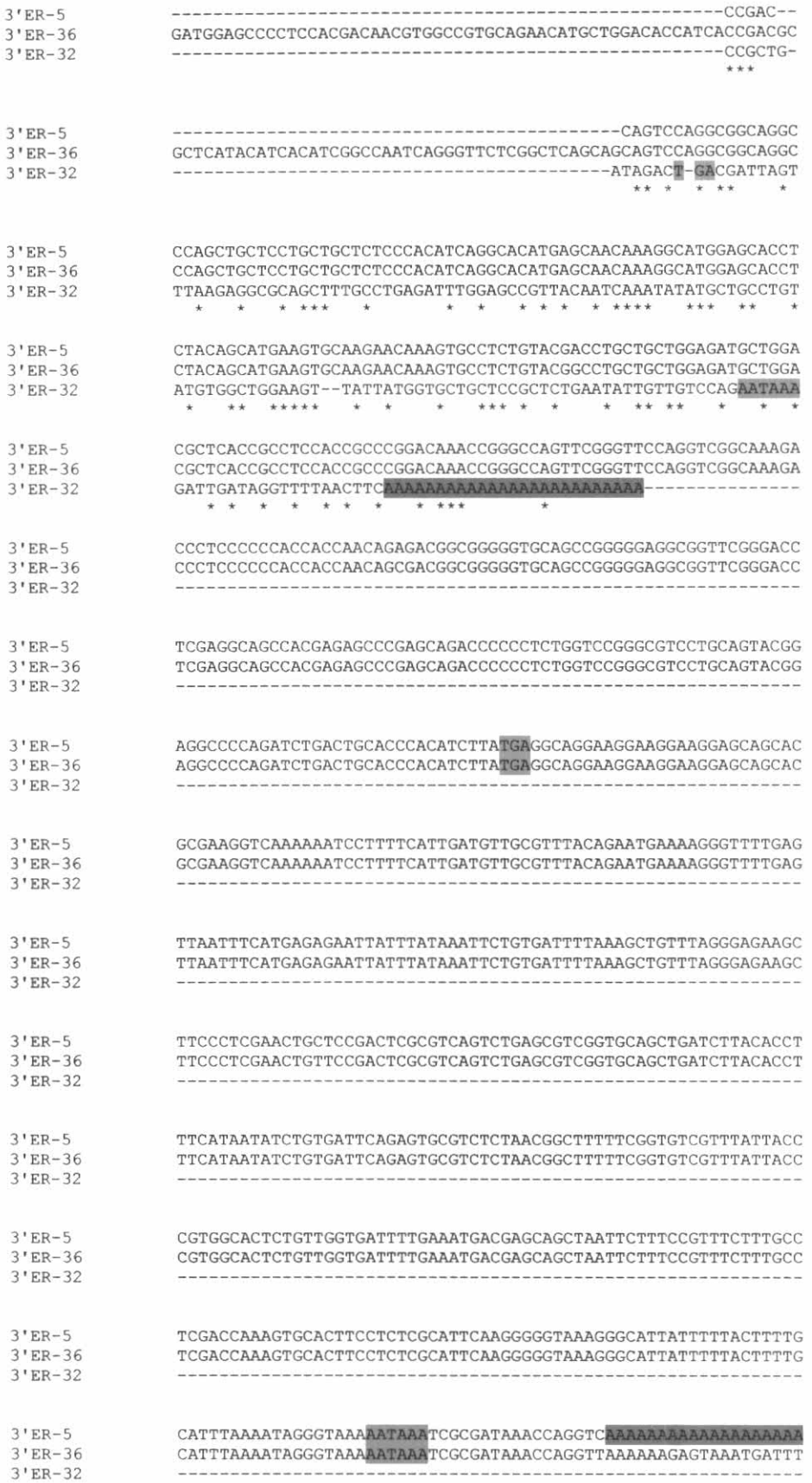


Figure 4.6 3'*ER* colony PCR product insert size = 1.76 kb digested with *Hin*dIII. Lane M is 100 bp DNA marker, other lanes is 3'*ER* colony PCR product insert size 1.76 digested with *Hin*dIII, No. of lane represent No. of clone.

Result in Fig. 4.6 showed 2 different patterns; group1 was No.13 and 36 and group2 was No. 22, 30, 42, 46, 49 and 51. No. 49 was sequenced (see in appendix B). BLAST analysis indicated it homologous with estrogen receptor alpha (*Chelon labrosus*) with E-value as 3e-160. Three types of 3'end cDNA sequence of ERα gene (3'*ER*/5, 32 and 36) were aligned as shown in Fig. 4.7

```
3'ER-5          CTTCAAGAGGAGCATCCAGGGTCACAATGACTACATGTGCCCGGCGACCAATCAGTGCAC
3'ER-36         CTTCAAGAGGAGCATCCAGGGTCACAATGACTACATGTGCCCGGCGACCAATCAGTGCAC
3'ER-32         CTTCAAGAGGAGCATCCAGGGTCACAATGACTACATGTGCCCGGCGACCAATCAGTGCAC
                ************************************************************

3'ER-5          TATTGACAGGAATCGGAGGAAGAGCTGCCAGGCTTGCCGTCTTAGGAAGTGTTATGAAGT
3'ER-36         TATTGACAGGAATCGGAGGAAGAGCTGCCAGGCTTGCCGTCTTAGGAAGTGTTATGAAGT
3'ER-32         TATTGACAGGAATCGGAGGAAGAGCTGCCAGGCTTGCCGTCTTAGGAAGTGTTATGAAGT
                ************************************************************

3'ER-5          GGGCATGATGAAAGGAGGTGTGCGCAAAGACCGGGGTCGCGTTTTGCGGCGCGACAAGCG
3'ER-36         GGGCATGATGAAAGGAGGTGTGCGCAAAGACCGGGGTCGCGTTTTGCGGCGCGACAAGCG
3'ER-32         GGGCATGATGAAAGGAGGTGTGCGCAAAGACCGGGGTCGCGTTTTGCGGCGCGACAAGCG
                ************************************************************

3'ER-5          GCGGGCCGGCGACAGAGACAAGGTTGCCAAAGATCCAGAGCACAAAGCGTCGGCGCCCCC
3'ER-36         GCGGGCCGGCGACAGAGACAAGGTTGCCAAAGATCCAGAGCACAAAGCGTCGGCGCCCCC
3'ER-32         GCGGGCCGGCGACAGAGACAAGGTTGCCAAAGATCCAGAGCACAAAGCGTCGGCGCCCCC
                ************************************************************

3'ER-5          TCAGGACGGGAGGAAGCGCAGCAGCAGCTGTAGCAGCACCGGAGGAGGAGGAGGAGGATC
3'ER-36         TCAGGACGGGAGGAAGCGCAGCAGCAGCTGTAGCAGCACCGGAGGAGGAGGAGGAGGATC
3'ER-32         TCAGGACGGGAGGAAGCGCAGCAGCAGCTGTAGCAGCACCGGAGGAGGAGGAG---GATC
                ************************************************    ****

3'ER-5          TTCTGTGACTAACTTGCCTCCAGACCAGGTGCTCCTCCTGCTCCAGGGAGCTGAGCCTCC
3'ER-36         TTCTGTGACTAACTTGCCTCCAGACCAGGTGCTCCTCCTGCTCCAGGGAGCTGAGCCTCC
3'ER-32         TTCTGTGACTAACTTGCCTCCAGACCAGGTGCTCCTCCTGCTCCAGGGAGCTGAGCCTCC
                ************************************************************

3'ER-5          GATGCTTTGCTCCCGGCAAAAGCTGAGCCGACCCTACACGGAGGTCACCATGATGACCCT
3'ER-36         GATGCTTTGCTCCCGGCAAAAGCTGAGCCGACCCTACACGGAGGTCACCATGATGACCCT
3'ER-32         GATGCTTTGCTCCCGGCAAAAGCTGAGCCGACCCTACACGGAGGTCACCATGATGACCCT
                ************************************************************

3'ER-5          GCTCACCAGCATGGCCGACAAGGAGCTGGTGCACATGATCGCCTGGGCCAAGAAGCTTCC
3'ER-36         GCTCACCAGCATGGCCGACAAGGAGCTGGTGCACATGATCGCCTGGGCCAAGAAGCTTCC
3'ER-32         GCTCACCAGCATGGCCGACAAGGAGCTGGTGCACATGATCGCCTGGGCCAAGAAGCTTCC
                ************************************************************

3'ER-5          AGGTTTCCTGCAGCTCTCCCTCCACGATCGGGTGCAGCTGCTGGAGAGCTCGTGGCTGGA
3'ER-36         AGGTTTCCTGCAGCTCTCCCTCCACGATCAGGTGCAGCTGCTGGAGAGCTCGTGGCTGGA
3'ER-32         AGGTTTGGAGAAGCAC-----CAACGA--GAGCATTTTTTCTGCA-ACCTTGTGTTTGAA
                ******   * *** *     * ****    *      * *** * * ** ***  ** *

3'ER-5          GGTGCTGATGATCGGGCTCATATGGAGGTCCACCCACTGTCCTGG---------------
3'ER-36         GGTGCTGATGATCGGGCTCATATGGAGGTCCACCCACTGTCCTGGAAAACTCATCTTCGC
3'ER-32         AAAACTG-CAGTGAAACAAATTTATTTGCCTCGACACGCCGCAGG---------------
                   ***    *    *  ** *    * *    ***    * **

3'ER-5          ------------------------------------GCAGTGGG-----------------
3'ER-36         CCAGGACCTCATACTGGACAGGAATGAAGGCGACTGCGTTGAGGGCATGGCTGAGATCTT
3'ER-32         ------------------------------------ACAGAAGAGAATCTGCAGG------
                                                    *

3'ER-5          ------------------------------------------------------------
3'ER-36         TGACATGCTCCTGGCCACCGCGTCCCGTTTCCGCATGCTCAAGCTCAAACCCGAGGAGTT
3'ER-32         ------------------------------------------------------------

3'ER-5          ------------------------------------------------------------
3'ER-36         CGTGTGTCTCAAAGCCATCATCCTGCTCAACTCTGGTGCCTTTTCTTTCTGCACCGGCAC
3'ER-32         ------------------------------------------------------------
```


3'ER-5 ----------CCGAC--
3'ER-36 GATGGAGCCCCTCCACGACAACGTGGCCGTGCAGAACATGCTGGACACCATCACCGACGC
3'ER-32 ----------CCGCTG-
3'ER-5 ----------CAGTCCAGGCGGCAGGC
3'ER-36 GCTCATACATCACATCGGCCAATCAGGGTTCTCGGCTCAGCAGCAGTCCAGGCGGCAGGC
3'ER-32 ----------ATAGACT-GACGATTAGT
3'ER-5 CCAGCTGCTCCTGCTGCTCTCCCACATCAGGCACATGAGCAACAAAGGCATGGAGCACCT
3'ER-36 CCAGCTGCTCCTGCTGCTCTCCCACATCAGGCACATGAGCAACAAAGGCATGGAGCACCT
3'ER-32 TTAAGAGGCGCAGCTTTGCCTGAGATTTGGAGCCGTTACAATCAAATATATGCTGCCTGT
3'ER-5 CTACAGCATGAAGTGCAAGAACAAAGTGCCTCTGTACGACCTGCTGCTGGAGATGCTGGA
3'ER-36 CTACAGCATGAAGTGCAAGAACAAAGTGCCTCTGTACGGCCTGCTGCTGGAGATGCTGGA
3'ER-32 ATGTGGCTGGAAGT--TATTATGGTGCTGCTCCGCTCTGAATATTGTTGTCCAGAATAAA
3'ER-5 CGCTCACCGCCTCCACCGCCCGGACAAACCGGGCCAGTTCGGGTTCCAGGTCGGCAAAGA
3'ER-36 CGCTCACCGCCTCCACCGCCCGGACAAACCGGGCCAGTTCGGGTTCCAGGTCGGCAAAGA
3'ER-32 GATTGATAGGTTTTAACTTCAAAAAAAAAAAAAAAAAAAAAAAA----------
3'ER-5 CCCTCCCCCCACCACCAACAGAGACGGCGGGGGTGCAGCCGGGGGAGGCGGTTCGGGACC
3'ER-36 CCCTCCCCCCACCACCAACAGCGACGGCGGGGGTGCAGCCGGGGGAGGCGGTTCGGGACC
3'ER-32 ----------
3'ER-5 TCGAGGCAGCCACGAGAGCCCGAGCAGACCCCCCTCTGGTCCGGGCGTCCTGCAGTACGG
3'ER-36 TCGAGGCAGCCACGAGAGCCCGAGCAGACCCCCCTCTGGTCCGGGCGTCCTGCAGTACGG
3'ER-32 ----------
3'ER-5 AGGCCCCAGATCTGACTGCACCCACATCTTATGAGGCAGGAAGGAAGGAAGGAGCAGCAC
3'ER-36 AGGCCCCAGATCTGACTGCACCCACATCTTATGAGGCAGGAAGGAAGGAAGGAGCAGCAC
3'ER-32 ----------
3'ER-5 GCGAAGGTCAAAAAATCCTTTTCATTGATGTTGCGTTTACAGAATGAAAAGGGTTTTGAG
3'ER-36 GCGAAGGTCAAAAAATCCTTTTCATTGATGTTGCGTTTACAGAATGAAAAGGGTTTTGAG
3'ER-32 ----------
3'ER-5 TTAATTTCATGAGAGAATTATTTATAAATTCTGTGATTTTAAAGCTGTTTAGGGAGAAGC
3'ER-36 TTAATTTCATGAGAGAATTATTTATAAATTCTGTGATTTTAAAGCTGTTTAGGGAGAAGC
3'ER-32 ----------
3'ER-5 TTCCCTCGAACTGCTCCGACTCGCGTCAGTCTGAGCGTCGGTGCAGCTGATCTTACACCT
3'ER-36 TTCCCTCGAACTGTTCCGACTCGCGTCAGTCTGAGCGTCGGTGCAGCTGATCTTACACCT
3'ER-32 ----------
3'ER-5 TTCATAATATCTGTGATTCAGAGTGCGTCTCTAACGGCTTTTTCGGTGTCGTTTATTACC
3'ER-36 TTCATAATATCTGTGATTCAGAGTGCGTCTCTAACGGCTTTTTCGGTGTCGTTTATTACC
3'ER-32 ----------
3'ER-5 CGTGGCACTCTGTTGGTGATTTTGAAATGACGAGCAGCTAATTCTTTCCGTTTCTTTGCC
3'ER-36 CGTGGCACTCTGTTGGTGATTTTGAAATGACGAGCAGCTAATTCTTTCCGTTTCTTTGCC
3'ER-32 ----------
3'ER-5 TCGACCAAAGTGCACTTCCTCTCGCATTCAAGGGGGTAAAGGGCATTATTTTTACTTTTG
3'ER-36 TCGACCAAAGTGCACTTCCTCTCGCATTCAAGGGGGTAAAGGGCATTATTTTTACTTTTG
3'ER-32 ----------
3'ER-5 CATTTAAAATAGGGTAAAAATAAATCGCGATAAACCAGGTCAAAAAAAAAAAAAAAAAAA
3'ER-36 CATTTAAAATAGGGTAAAAATAAATCGCGATAAACCAGGTTAAAAAAGAGTAAATGATTT
3'ER-32 ----------

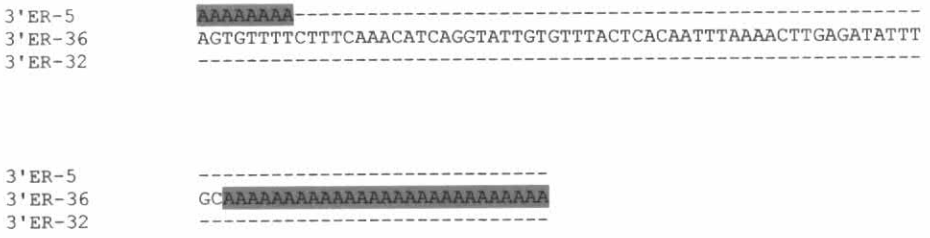


Figure 4.7 Nucleotide alignment of 3 types of 3'end of cDNA sequences of ERα gene (3'*ER*/5, 32 and 36) in *L. subviridis*. Yellow, blue, pink and green label are *ER* F-RACE primer, stop codon, polyadenylation signal and polyA tail, respectively.

Nucleotide sequences of 3 variant transcript of ERα gene were different. 3'*ER*/5 and 3'*ER*/32 lack some region of coding sequences but 3'*ER*/36 have complete coding sequence. Three variant of ERα gene contained stop codon, 1 polyadenylation signal (AATAAA), and polyA tail which different in length and polyadenylation site between variants. 3'untranslated region (3'UTR) were different in length between variants as shown in Fig. 4.7. Three types of 3'end cDNA sequence of ERα gene (3'*ER*/5, 32 and 36) were translated to amino acid sequences and aligned. DNA binding domain showed 100 % identical in 3 variant. DNA binding domain contained zinc finger region (red alphabet shown in Fig. 4.8) and D box (PATNQ) (underline in Fig. 4.8 which involved in receptor dimerization. D domain showed 100 % identical in variant 5 and 36, 99.08 % similarity between variant (5 and 36) with variant 32. Interestingly 1 amino acid residues is missed from variant 32. Ligand binding domain was complete in variant 36 but lack 66.67 % of this domain in variant 5 and 32. F domain was complete in variant 36 and 5 with 98.97 % similarity but lack almost portion of this domain in variant 32 as shown in Fig. 4.8.

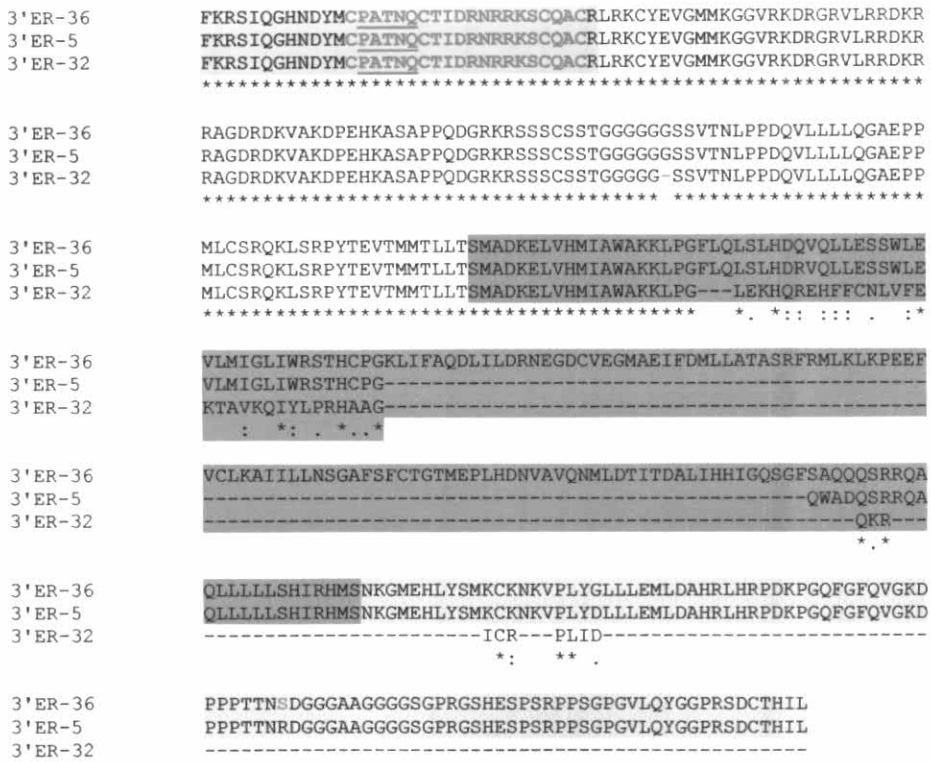


Figure 4.8 Deduced sequence of 3'*ER*/5, 32 and 36. Yellow label, red alphabet, underline, non label, blue label, and grey label is DNA binding domain, zinc finger region, D box, D domain, ligand binding domain, and F domain, respectively.

Nucleotide sequences of 3'*ER*/36 and 49 which were similar to *Hin*dIII digested pattern were aligned and showed in Fig. 4.9 A and B

```
3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           CTTCAAGAGGAGCATCCAGGGTCACAATGACTACATGTGCCCGGCGACCAATCAGTGCAC

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           TATTGACAGGAATCGGAGGAAGAGCTGCCAGGCTTGCCGTCTTAGGAAGTGTTATGAAGT

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           GGGCATGATGAAAGGAGGTGTGCGCAAAGACCGGGGTCGCGTTTTGCGGCGCGACAAGCG

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           GCGGGCCGGCGACAGAGACAAGGTTGCCAAAGATCCAGAGCACAAAGCGTCGGCGCCCCC

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           TCAGGACGGGAGGAAGCGCAGCAGCAGCTGTAGCAGCACCGGAGGAGGAGGAGGAGGATC

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           TTCTGTGACTAACTTGCCTCCAGACCAGGTGCTCCTCCTGCTCCAGGGAGCTGAGCCTCC

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           GATGCTTTGCTCCCGGCAAAAGCTGAGCCGACCCTACACGGAGGTCACCATGATGACCCT

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           GCTCACCAGCATGGCCGACAAGGAGCTGGTGCACATGATCGCCTGGGCCAAGAAGCTTCC

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           AGGTTTCCTGCAGCTCTCCCTCCACGATCAGGTGCAGCTGCTGGAGAGCTCGTGGCTGGA

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           GGTGCTGATGATCGGGCTCATATGGAGGTCCACCCACTGTCCTGGAAAACTCATCTTCGC

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           CCAGGACCTCATACTGGACAGGAATGAAGGCGACTGCGTTGAGGGCATGGCTGAGATCTT

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           TGACATGCTCCTGGCCACCGCGTCCCGTTTCCGCATGCTCAAGCTCAAACCCGAGGAGTT


3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           CGTGTGTCTCAAAGCCATCATCCTGCTCAACTCTGGTGCCTTTTCTTTCTGCACCGGCAC

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           GATGGAGCCCCTCCACGACAACGTGGCCGTGCAGAACATGCTGGACACCATCACCGACGC

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           GCTCATACATCACATCGGCCAATCAGGGTTCTCGGCTCAGCAGCAGTCCAGGCGGCAGGC

3'ER-49-M13F      ------------------------------------------------------------
3'ER-36           CCAGCTGCTCCTGCTGCTCTCCCACATCAGGCACATGAGCAACAAAGGCATGGAGCACCT


3'ER-49-M13F      ------------------------------------------------------CTGGA
3'ER-36           CTACAGCATGAAGTGCAAGAACAAAGTGCCTCTGTACGGCCTGCTGCTGGAGATGCTGGA
                                                                     *****

3'ER-49-M13F      CGCTCACCGCCTCCACCGCCCGGACAAACCGGGCCAGTTCGGGTTCCAGGTCGGCAAAGA
3'ER-36           CGCTCACCGCCTCCACCGCCCGGACAAACCGGGCCAGTTCGGGTTCCAGGTCGGCAAAGA
                  ************************************************************

3'ER-49-M13F      CCCTCCCCCCACCCACCAACAGCGACGGCGGGGGTGCAGCCGGGGGAGGCGGTTCGGGAC
3'ER-36           CCCTCCCCCCACC-ACCAACAGCGACGGCGGGGGTGCAGCCGGGGGAGGCGGTTCGGGAC
                  ************ ************************************************
```

```
3'ER-49-M13F      CTCGAGGCAGCCACGAGAGCCCGAGCAGACCCCCCTCTGGTCCGGGCGTCCTGCAGTACG
3'ER-36           CTCGAGGCAGCCACGAGAGCCCGAGCAGACCCCCCTCTGGTCCGGGCGTCCTGCAGTACG
                  ************************************************************

3'ER-49-M13F      GAGGCCCCAGATCTGACTGCACCCACATCTTATGAGGCAGGAAGGAAGGAAGGAGCAGCA
3'ER-36           GAGGCCCCAGATCTGACTGCACCCACATCTTATGAGGCAGGAAGGAAGGAAGGAGCAGCA
                  ************************************************************

3'ER-49-M13F      CGCGAAGGTCAAAAAAATCCTTTTCATTGATGTTGCGTTTACAGAATGAAAAGGGTTTTG
3'ER-36           CGCGAAGGTCAAAAAA-TCCTTTTCATTGATGTTGCGTTTACAGAATGAAAAGGGTTTTG
                  ********** ************************************************

3'ER-49-M13F      AGTCAATTTCATGAGAGAATTATTTATAAATTCTGTGATTTTAAAGCTGTTTAGGGAGAA
3'ER-36           AGTTAATTTCATGAGAGAATTATTTATAAATTCTGTGATTTTAAAGCTGTTTAGGGAGAA
                  *** ********************************************************

3'ER-49-M13F      GCTTCCCTCGAACTGCTCCGACTCGCGTCAGTTTGAGCGTCGGTGCAGCTGATCTTACAC
3'ER-36           GCTTCCCTCGAACTGTTCCGACTCGCGTCAGTCTGAGCGTCGGTGCAGCTGATCTTACAC
                  ********** ********** ************************************

3'ER-49-M13F      CTTTCATAATATCTGTGATTCAGAGTGCGTCTCTAACGGCTTTTTCGGTGTCGTTTATTA
3'ER-36           CTTTCATAATATCTGTGATTCAGAGTGCGTCTCTAACGGCTTTTTCGGTGTCGTTTATTA
                  ************************************************************

3'ER-49-M13F      CCCGTGGCACTCTGTTGGTGATTTTGAAATGACGAGCAGCTAATTCTTTCCGTTTTTTTG
3'ER-36           CCCGTGGCACTCTGTTGGTGATTTTGAAATGACGAGCAGCTAATTCTTTCCGTTTCTTTG
                  ********************************************************** ****

3'ER-49-M13F      CCTCGACCAAAGTGCACTTCCTCTCGCATTCAAGGGGGTAAAGGGCATTATTTTTACTTT
3'ER-36           CCTCGACCAAAGTGCACTTCCTCTCGCATTCAAGGGGGTAAAGGGCATTATTTTTACTTT
                  ************************************************************

3'ER-49-M13F      TGCATTTAAAATAGGGTAAAAATAAATCGCGATAAACCAGGTTAAAAAAGAGTAAATGAT
3'ER-36           TGCATTTAAAATAGGGTAAAAATAAATCGCGATAAACCAGGTTAAAAAAGAGTAAATGAT
                  ************************************************************

3'ER-49-M13F      TTAGTGTTTTCTTTCAAACATCAGGTATTGTGTTTACTCACAATTTAAAACTTGAGATAT
3'ER-36           TTAGTGTTTTCTTTCAAACATCAGGTATTGTGTTTACTCACAATTTAAAACTTGAGATAT
                  ************************************************************

3'ER-49-M13F      TTGGAAAAAAAAAAAAAAAAAAAAAAAAAAA
3'ER-36           TTGCAAAAAAAAAAAAAAAAAAAAAAAAAA-
                  *** ****************************
```

A

```
3'ER-36           CTTCAAGAGGAGCATCCAGGGTCACAATGACTACATGTGCCCGGCGACCAATCAGTGCAC
3'ER-49-M13R      CTTCAAGAGGAGCATCCAGGGTCACAATGACTACATGTGCCCGGCGACCAATCAGTGCAC
                  ************************************************************

3'ER-36           TATTGACAGGAATCGGAGGAAGAGCTGCCAGGCTTGCCGTCTTAGGAAGTGTTATGAAGT
3'ER-49-M13R      TATTGACAGGAATCGGAGGAAGAGCTGCCAGGCTTGCCGTCTTAGGAAGTGTTATGAAGT
                  ************************************************************

3'ER-36           GGGCATGATGAAAGGAGGTGTGCGCAAAGACCGGGGTCGCGTTTTGCGGCGCGACAAGCG
3'ER-49-M13R      GGGCATGATGAAAGGAGGTGTGCGCAAAGACCGGGGTCGCGTTTTGCGGCGCGACAAGCG
                  ************************************************************

3'ER-36           GCGGGCCGGCGACAGAGACAAGGTTGCCAAAGATCCAGAGCACAAAGCGTCGGCGCCCCC
3'ER-49-M13R      GCGGGCCGGCGACAGAGACAAGGTTGCCAAAGATCCAGAGCACAAAGCGTCGGCGCCCCC
                  ************************************************************

3'ER-36           TCAGGACGGGAGGAAGCGCAGCAGCAGCTGTAGCAGCACCGGAGGAGGAGGAGGAGGA--
3'ER-49-M13R      TCAGGACGGGAGGAAGCGCAGCAGCAGCTGTAGCAGCACCGGAGGAGGAGGAGGAGGAGG
                  ************************************************************

3'ER-36           -TCTTCTGTGACTAACTTGCCTCCAGACCAGGTGCTCCTCCTGCTCCAGGGAGCTGAGCC
3'ER-49-M13R      ATCTTCTGTGACTAACTTGCCTCCAGACCAGGTGCTCCTCCTGCTCCAGGGAGCTGAGCC
                   ***********************************************************

3'ER-36           TCCGATGCTTTGCTCCCGGCAAAAGCTGAGCCGACCCTACACGGAGGTCACCATGATGAC
3'ER-49-M13R      TCCGATGCTTTGCTCCCGGCAAAAGCTGAGCCGACCCTACACGGAGGTCACCATGATGAC
                  ************************************************************
```

```
3'ER-36          CCTGCTCACCAGCATGGCCGACAAGGAGCTGGTGCACATGATCGCCTGGGCCAAGAAGCT
3'ER-49-M13R     CCTGCTCACCAGCATAGCCGACAAGGAGCTGGTGCACATGATCGCCTGGGCCAAGAAGCT
                 ********** ********************************************

3'ER-36          TCCAGGTTTCCTGCAGCTCTCCCTCCACGATCAGGTGCAGCTGCTGGAGAGCTCGTGGCT
3'ER-49-M13R     TCCAGGTTTCCTGCAGCTCTCCCTCCACGATCAGGTGCAGCTGCTGGAGAGCTCGTGGCT
                 ************************************************************

3'ER-36          GGAGGTGCTGATGATCGGGCTCATATGGAGGTCCACCCACTGTCCTGGAAAACTCATCTT
3'ER-49-M13R     GGAGGTGCTGATGATCGGGCTCATATGGAGGTCCACCCACTGTCCTGGAAAACTCATCTT
                 ************************************************************

3'ER-36          CGCCCAGGACCTCATACTGGACAGGAATGAAGGCGACTGCGTTGAGGGCATGGCTGAGAT
3'ER-49-M13R     CGCCCAGGACCTCATACTGGACAGGAATGAAGGCGACTGCGTTGAGGGCATGGCTGAGAT
                 ************************************************************

3'ER-36          CTTTGACATGCTCCTGGCCACCGCGTCCCGTTTCCGCATGCTCAAGCTCAAACCCGAGGA
3'ER-49-M13R     CTTTGACATGCTCCTGGCCACCGCGTCCCGTTTCCGCATGCTCAAGCTCAAACCCGAGGA
                 ************************************************************

3'ER-36          GTTCGTGTGTCTCAAAGCCATCATCCTGCTCAACTCTGGTGCCTTTTCTTTCTGCACCGG
3'ER-49-M13R     GTTCGTGTGTCTC-------------------------------------------------
                 *************

3'ER-36          CACGATGGAGCCCCTCCACGACAACGTGGCCGTGCAGAACATGCTGGACACCATCACCGA
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          CGCGCTCATACATCACATCGGCCAATCAGGGTTCTCGGCTCAGCAGCAGTCCAGGCGGCA
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          GGCCCAGCTGCTCCTGCTGCTCTCCCACATCAGGCACATGAGCAACAAAGGCATGGAGCA
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          CCTCTACAGCATGAAGTGCAAGAACAAAGTGCCTCTGTACGGCCTGCTGCTGGAGATGCT
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          GGACGCTCACCGCCTCCACCGCCCGGACAAACCGGGCCAGTTCGGGTTCCAGGTCGGCAA
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          AGACCCTCCCCCCACCACCAACAGCGACGGCGGGGGTGCAGCCGGGGGAGGCGGTTCGGG
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          ACCTCGAGGCAGCCACGAGAGCCCGAGCAGACCCCCCTCTGGTCCGGGCGTCCTGCAGTA
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          CGGAGGCCCCAGATCTGACTGCACCCACATCTTATGAGGCAGGAAGGAAGGAAGGAGCAG
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          CACGCGAAGGTCAAAAAATCCTTTTCATTGATGTTGCGTTTACAGAATGAAAAGGGTTTT
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          GAGTTAATTTCATGAGAGAATTATTTATAAATTCTGTGATTTTAAAGCTGTTTAGGGAGA
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          AGCTTCCCTCGAACTGTTCCGACTCGCGTCAGTCTGAGCGTCGGTGCAGCTGATCTTACA
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          CCTTTCATAATATCTGTGATTCAGAGTGCGTCTCTAACGGCTTTTTCGGTGTCGTTTATT
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          ACCCGTGGCACTCTGTTGGTGATTTTGAAATGACGAGCAGCTAATTCTTTCCGTTTCTTT
3'ER-49-M13R     ------------------------------------------------------------
```

```
3'ER-36          GCCTCGACCAAAGTGCACTTCCTCTCGCATTCAAGGGGGTAAAGGGCATTATTTTTACTT
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          TTGCATTTAAAATAGGGTAAAAATAAATCGCGATAAACCAGGTTAAAAAAGAGTAAATGA
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          TTTAGTGTTTTCTTTCAAACATCAGGTATTGTGTTTACTCACAATTTAAAACTTGAGATA
3'ER-49-M13R     ------------------------------------------------------------

3'ER-36          TTTGCAAAAAAAAAAAAAAAAAAAAAAAAA
3'ER-49-M13R     ------------------------------
```

B

Figure 4.9 3'*ER*/36 aligned with 3'*ER*/49 M13F sequence (A) and M13R (B). Blue, pink and green label were stop codon, polyadenylation signal and polyA tail, respectively.

cDNA sequences shown 98.57 % similarity between 3'*ER*/49 and 3'*ER*/36 as shown in Fig. 4.9 so the internal sequence of 3'*ER*/49 was not further sequenced.

*ER*α R RACE primer was designed from 3' cDNA sequence of ERα gene. A DNA fragment at the size of 1.2 kb was obtained from 5'RACE PCR amplification of *ER*α in liver (Fig.4.10). After sequencing (see in appendix B) and BLAST analysis (see in appendix C), this PCR product was homologous with estrogen receptor alpha of Nile tilapia (*Oreochromis niloticus niloticus*) with E-value of 0.0

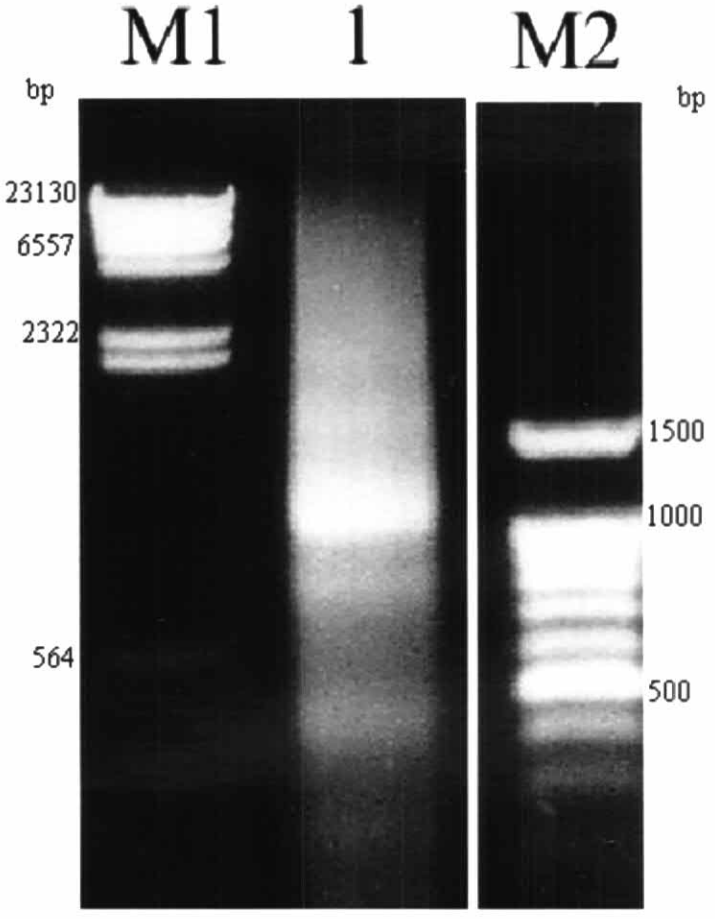


Figure 4.10 5'RACE PCR product of *ERα* separated in 1.2% agarose gel. Lane M1 is λ*Hin*dIII, M2 is 100 bp DNA marker and lane 1 is 5'RACE PCR product of *ERα*.

4.3.3 Full length cDNA sequence of *ERα*

5'*ERα* 1.2/37 and 3'*ER*/36 cDNA sequences were ligated *in silico* and full length cDNA sequence of *ERα* was obtained. The sequence was translated to deduced amino acid sequence of ERα (Fig 4.11).

```
gtgtccaatgagcccagctcgatgcctgtgaggtgagcggcagggacagcgagggatcaa
actgtgtctgtgtctgtgtgtgtgtcgcccagcagattaaggggggatgattcatctat
                                                          M  Y
aagaggcagagcccagcgcagagcagacagccgtgcggaccagtcctcaaacccaggaac
 K  R  Q  S  P  A  Q  S  R  Q  P  C  G  P  V  L  K  P  R  N
agcccagcttcctcagagctggagaccctgtccccttcgcgcccctctccttcaccgcgc
 S  P  A  S  S  E  L  E  T  L  S  P  S  R  P  S  P  S  P  R
gccccccctcggtgacatgtaccctgaagagagccgggggtctgcagggacagccactgtg
 A  P  L  G  D  M  Y  P  E  E  S  R  G  S  A  G  T  A  T  V
gacttcctggaagggacgtacgactatgccgccccccagccctgccccgactcctctttac
 D  F  L  E  G  T  Y  D  Y  A  A  P  S  P  A  P  T  P  L  Y
agccactccaccactggctactattccgctcctctggacgctcacggaccaccctctgat
 S  H  S  T  T  G  Y  Y  S  A  P  L  D  A  H  G  P  P  S  D
ggcagccttcagtccctgggcagtgggccgaccagtcctcttgtgtttgtgccctccagc
 G  S  L  Q  S  L  G  S  G  P  T  S  P  L  V  F  V  P  S  S
ccacggctcagccctttttatgcacccacccagccaccactatctggaaaccacctcgaca
 P  R  L  S  P  F  M  H  P  P  S  H  H  Y  L  E  T  T  S  T
cccgtctacagatccagccaccagccgctgtccagagaggagcagcagtgtggcgcccgc
 P  V  Y  R  S  S  H  Q  P  L  S  R  E  E  Q  Q  C  G  A  R
gacgaggcgtacgcttcgggggagctgggagctgggccgggggcttcgagatgaccaag
 D  E  A  Y  A  S  G  E  L  G  A  G  A  G  G  F  E  M  T  K
gagactcgtttctgtgccgtctgcagcgactacgcctctgggtaccactacggggtgtgg
 E  T  R  F  C  A  V  C  S  D  Y  A  S  G  Y  H  Y  G  V  W
tcctgtgagggctgcaaggccttcttcaagaggagcatccagggtcacaatgactacatg
 S  C  E  G  C  K  A  F  F  K  R  S  I  Q  G  H  N  D  Y  M
tgcccggcgaccaatcagtgcactattgacaggaatcggaggaagagctgccaggcttgc
 C  P  A  T  N  Q  C  T  I  D  R  N  R  R  K  S  C  Q  A  C
cgtcttaggaagtgttatgaagtgggcatgatgaaaggaggtgtgcgcaaagaccggggt
 R  L  R  K  C  Y  E  V  G  M  M  K  G  G  V  R  K  D  R  G
cgcgttttgcggcgcgacaagcggcgggccggcgacagagacaaggttgccaaagatcca
 R  V  L  R  R  D  K  R  R  A  G  D  R  D  K  V  A  K  D  P
gagcacaaagcgtcggcgccccctcaggacgggaggaagcgcagcagcagctgtagcagc
 E  H  K  A  S  A  P  P  Q  D  G  R  K  R  S  S  S  C  S  S
accggaggaggaggaggaggatcttctgtgactaacttgcctccagaccaggtgctcctc
 T  G  G  G  G  G  G  S  S  V  T  N  L  P  P  D  Q  V  L  L
ctgctccagggagctgagcctccgatgctttgctcccggcaaaagctgagccgaccctac
 L  L  Q  G  A  E  P  P  M  L  C  S  R  Q  K  L  S  R  P  Y
acggaggtcaccatgatgaccctgctcaccagcatggccgacaaggagctggtgcacatg
 T  E  V  T  M  M  M  T  L  L  T  S  M  A  D  K  E  L  V  H  M
atcgcctgggccaagaagcttccaggtttcctgcagctctccctccacgatcaggtgcag
 I  A  W  A  K  K  L  P  G  F  L  Q  L  S  L  H  D  Q  V  Q
ctgctggagagctcgtggctggaggtgctgatgatcgggctcatatggaggtccacccac
 L  L  E  S  S  W  L  E  V  L  M  I  G  L  I  W  R  S  T  H
tgtcctggaaaactcatcttcgcccaggacctcatactggacaggaatgaaggcgactgc
 C  P  G  K  L  I  F  A  Q  D  L  I  L  D  R  N  E  G  D  C
gttgagggcatggctgagatctttgacatgctcctggccaccgcgtcccgtttccgcatg
 V  E  G  M  A  E  I  F  D  M  L  L  A  T  A  S  R  F  R  M
ctcaagctcaaacccgaggagttcgtgtgtctcaaagccatcatcctgctcaactctggt
 L  K  L  K  P  E  E  F  V  C  L  K  A  I  I  L  L  N  S  G
gccttttctttctgcaccggcacgatggagcccctccacgacaacgtggccgtgcagaac
 A  F  S  F  C  T  G  T  M  E  P  L  H  D  N  V  A  V  Q  N
atgctggacaccatcaccgacgcgctcatacatcacatcggccaatcagggttctcggct
 M  L  D  T  I  T  D  A  L  I  H  H  I  G  Q  S  G  F  S  A
cagcagcagtccaggcggcaggcccagctgctcctgctgctctcccacatcaggcacatg
 Q  Q  Q  S  R  R  Q  A  Q  L  L  L  L  L  S  H  I  R  H  M
agcaacaaaggcatggagcacctctacagcatgaagtgcaagaacaaagtgcctctgtac
 S  N  K  G  M  E  H  L  Y  S  M  K  C  K  N  K  V  P  L  Y
ggcctgctgctggagatgctggacgctcaccgcctccaccgcccggacaaaccgggccag
 G  L  L  L  E  M  L  D  A  H  R  L  H  R  P  D  K  P  G  Q
ttcgggttccaggtcggcaaagaccctccccccaccaccaacagcgacggcgggggtgca
 F  G  F  Q  V  G  K  D  P  P  P  T  T  N  S  D  G  G  G  A
gccgggggaggcggttcgggacctcgaggcagccacgagagcccgagcagaccccctct
 A  G  G  G  G  S  G  P  R  G  S  H  E  S  P  S  R  P  P  S
ggtccgggcgtcctgcagtacggaggccccagatctgactgcacccacatcttatgaggc
  G  P  G  V  L  Q  Y  G  G  P  R  S  D  C  T  H  I  L
```

aggaaggaaggaaggagcagcacgcgaaggtcaaaaaatccttttcattgatgttgcgtt
tacagaatgaaaagggttttgagttaatttcatgagagaattatttataaattctgtgat
tttaaagctgtttagggagaagcttccctcgaactgttccgactcgcgtcagtctgagcg
tcggtgcagctgatcttacacctttcataatatctgtgattcagagtgcgtctctaacgg
ctttttcggtgtcgtttattacccgtggcactctgttggtgattttgaaatgacgagcag
ctaattctttccgtttctttgcctcgaccaaagtgcacttcctctcgcattcaaggggt
aaagggcattatttttactttgcatttaaaatagggtaaaaataaatcgcgataaacca
ggttaaaaaagagtaaatgatttagtgttttctttcaaacatcaggtattgtgtttactc
acaatttaaaacttgagatatttgcaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

Figure 4.11 Full length cDNA sequence of *ERα* and deduced amino acid sequence of ERα. Blue, red alphabet, blue alphabet, pink and green labels are *ERα* full length primer, start codon, stop codon, polyadenylation signal and polyA tail, respectively.

Result in Fig. 4.11 shows full length mRNA sequence of *L. subviridis ERα* at the size of 2,512 bp which contained 5'UTR, ORF and 3'UTR (contained 1 polyadenylation site) at the sizes of 114, 1,863 and 535 bp, respectively. ORF at the size of 1,863 bp encoded deduced amino acid sequence of ERα which contained 620 amino acid residues with the molecular weight at 67.55 kDa. ERα deduced sequence was analyzed with Prosite and SMART program. Domains and motifs found in the sequence showed in Fig. 4.12 and table 4.1 AF1-containing modulatory domain was found at the N-terminus in estrogen alpha-type receptors at amino acid residues position 7-183 (E-value = 1.10e-04), Nuclear hormone receptors DNA-binding domain found at amino acid residues position 184-243 (score = 19.144), zinc finger region (NR C4-type) found at amino acid residues position 187-207 and 223-242, P box directly interact with regulatory region of DNA found at amino acid residues position 205-209, D box involved in receptor dimerization found at amino acid residues position 224-228, and ligand binding domain of hormone receptors (E-value = 1.94e-33) were found at amino acid residues position 353-524.

Table 4.1 Domains, motifs, and consensus patterns of ERα

Domain	Site	Consensus pattern
AF1-containing modulatory domain	7-183	Poorly conserved region which spans the first 180 residues and contains the activation function 1 (AF1) region (http://pfam.janelia.org/cgi-bin/getdesc?name=Oest_recep)
DNA-binding domain	184-243	C - x(2) - C - x(1,2) - [DENAVSPHKQT] - x(5,6) - [HNY] - [FY] - x(4) - C - x(2) - C - x(2) - F(2) - x – R The 4 C's are zinc ligands (http://br.expasy.org/prosite)
P box	205-209	EGCKA (Attwood, Kroll, and Denslow, 2004)
D box	224-228	PATNQ (Attwood, Kroll, and Denslow, 2004)
zinc finger region (NR C4-type)	187-207 , 223-242	C - x(2) - C - x(1,2) - [DENAVSPHKQT] - x(5,6) - [HNY] - [FY] - x(4) - C - x(2) - C - x(2) - F(2) - x - R (http://br.expasy.org/prosite)
ligand binding domain	353-523	Contained helices H3, H6, H8, and H11 surrounding the hydrophobic cavity in tertiary structure (Menuet et al, 2002)

MYKRQSPAQSRQPCGPVLKPRNSPASSELETLSPSRPSPSPRAPLGDMYPEESRGSAGTATVDFLEGTYDYAAPSPAPTPLYSH
STTGYYSAPLDAHGPPSDGSLQSLGSGPTSPLVFVPSSPRLSPFMHPPSHHYLETTSTPVYRSSHQPLSREEQQCGARDEAYAS
GELGAGAGGFEMTKETRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDRNRRKSCQACRLRKCYEVGM
MKGGVRKDRGRVLRRDKRRAGDRDKVAKDPEHKASAPPQDGRKRSSSCSSTGGGGGGSSVTNLPPDQVLLLLQGAEPPMLCSRQ
KLSRPYTEVTMMTLLTSMADKELVHMIAWAKKLPGFLQLSLHDQVQLLESSWLEVLMIGLIWRSTHCPGKLIFAQDLILDRNEG
DCVEGMAEIFDMLLATASRFRMLKLKPEEFVCLKAIILLNSGAFSFCTGTMEPLHDNVAVQNMLDTITDALIHHIGQSGFSAQQ
QSRRQAQLLLLLSHIRHMSNKGMEHLYSMKCKNKVPLYGLLLEMLDAHRLHRPDKPGQFGFQVGKDPPPTTNSDGGGAAGGGGS
GPRGSHESPSRPPSGPGVLQYGGPRSDCTHIL

Figure 4.12 ERα of *L. subviridis* showed important domains and motifs. Green label is AF1-containing modulatory domain, yellow label is nuclear hormone receptors DNA-binding domain, underline is zinc finger region (NR C4-type), blue alphabet is P box, red alphabet is D box and blue label is ligand binding domain of hormone receptors. ERα deduced sequence aligned with ERα amino acid sequences of several fish species was shown in Fig. 4.13

```
Oreochromis_niloticus_niloticus   ------------------------------------------------MYP
Astatotilapia_burtoni             ------------------------------------------------MYP
Liza_subviridis                   MYKRQSPAQSRQPCGPVLKPRNSPASSELETLSPSRPSPSPRAPLGDMYP
Zoarces_viviparus                 ------------------------------------------------MYP
Acanthopagrus_schlegelii          ------------------------------------------------MYP
Halichoeres_tenuispinis           ------------------------------------------------MYP
Ictalurus_punctatus               --MSEEQARAEAPAGARQRRRSELEG---------YSVSLASLKLSPMYP
                                                                                 ***

Oreochromis_niloticus_niloticus   EE-SRGSGGVA-TVDFLEGTYDYAAPTPAPTPLYSHSTTGCYSAPLDAHG
Astatotilapia_burtoni             EE-SRGSGGVA-TVDFLEGSYDYAAPTPAPTPLYSHSTTGCYSAPLDAHG
Liza_subviridis                   EE-SRGSAGTA-TVDFLEGTYDYAAPSPAPTPLYSHSTTGYYSAPLDAHG
Zoarces_viviparus                 EE-SRGSGGVT-TVDFLEGTYDYAAPTPAPTPLYSHSTPGFYSAPLDSHR
Acanthopagrus_schlegelii          ED-SRVSGGVA-TVDFLEGTYDYAAPTPAPTPLYSHSTPGYYSAPLDAHG
Halichoeres_tenuispinis           EE-SRGSGGVG-TVDFLEGTYDYTAPTPAPT-LYSLSTQGYYSAALDTHG
Ictalurus_punctatus               EEEQRTTGGISSTAHYLDGTFNYTTNPDATN-----SSVDYYSVAPE---
                                  *: .* :.*    *..:*:*:::*:: . *..     *: . **.. :

Oreochromis_niloticus_niloticus   PLSDGSLQSLGSGPTSPLVFVPSSPRLSPFM-HPPS--------HHYLET
Astatotilapia_burtoni             PPSDGSLQSLGSGTTSPLVFVPSSPRLSPFM-HPPS--------HHYLET
Liza_subviridis                   PPSDGSLQSLGSGPTSPLVFVPSSPRLSPFM-HPPS--------HHYLET
Zoarces_viviparus                 PPSDGSLQSLGSGPTSPIVFVPSSPRLSPFL-HPPS--------HHYLET
Acanthopagrus_schlegelii          PPSDGSLQSLGSGPNSPLVFVPSSPRLSPFM-HPPT--------HHYLET
Halichoeres_tenuispinis           QPSDSSIQSLGSGPSSPLVFVPSSPRLSPFM-HLPS--------HHYLET
Ictalurus_punctatus               -PQEENLQPLPNGSSSPPVFVPSSPQLSPFLGHPPAGQHTAQQVPYYLEP
                                   .: .:*.* .*..** *******:****: * *:          :***.

Oreochromis_niloticus_niloticus   TSTPVY----RSSHQPVPREDQ-CGTRDEAYSVGELG---AGAG--GFEM
Astatotilapia_burtoni             TSTPVY----RSSHQPVPRDDQ-CGTRDEAYGLGELG---AGAG--GFEM
Liza_subviridis                   TSTPVY----RSSHQPLSREEQQCGARDEAYASGELG---AGAG--GFEM
Zoarces_viviparus                 TSTHVYS--VPSSQQSVSREDQ-CGTSDESYSLVESG---AGAGAGGFEM
Acanthopagrus_schlegelii          TSTPIYRSSVPSSQHSASREDQ-CGTSDDSYSVGESG---AGAGAAGFEM
Halichoeres_tenuispinis           SSTPVYRSSVSSSQQSISREEH-CGTSDESYSMGESG---AGAAAGCFEM
Ictalurus_punctatus               SGTSIYRSSVLASAGSRVELCSAPGRQDVYTAVGASGPSGASGPSGAIGL
                                  :.* :*      :*  .  .     *  *   .    *   *..    : :

Oreochromis_niloticus_niloticus   TKDTRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCT
Astatotilapia_burtoni             TKETRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCT
Liza_subviridis                   TKETRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCT
Zoarces_viviparus                 AKETRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCT
Acanthopagrus_schlegelii          AKEMRFCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCT
Halichoeres_tenuispinis           AKEMRYCAVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCT
Ictalurus_punctatus               VKEIRYCSVCSDYASGYHYGVWSCEGCKAFFKRSIQGHNDYVCPATNQCT
                                  .*: *:*:*********************************:********

Oreochromis_niloticus_niloticus   IDKNRRKSCQACRLRKCYEVGMMKGGMRKDRG-RVLRREKRR--ACDRDK
Astatotilapia_burtoni             IDKNRRKSCQACRLRKCYEVGMMKGGMRKDRG-RVLRREKRR--AYDRDK
Liza_subviridis                   IDRNRRKSCQACRLRKCYEVGMMKGGVRKDRG-RVLRRDKRR--AGDRDK
Zoarces_viviparus                 IDRNRRKSCQACRLRKCYEVGMMKGGVRKDRG-HVLRRDKRRSDLSDRDK
Acanthopagrus_schlegelii          IDRNRRKSCQACRLRKCYEVGMMKGGVRKDRG-RVLRRDKRRTGTSDRDK
Halichoeres_tenuispinis           IDRNRRKSCQACRLRKCYEVGMMKGGVRKDRG-RVLRRDKRRTGTSDKDN
Ictalurus_punctatus               IDRNRRKSCQACRLRKCYEVGMMKGGFRKERGGRIIKHNRRPSGLKERER
                                  **:***********************.**:** :::::::*      :::.

Oreochromis_niloticus_niloticus   ---PAKDLPHTR-ASPQDGRKRAMSSSSTSGGGGRSSLNNMPPDQVLLLL
Astatotilapia_burtoni             ---PAKDLPHTK-APPHDGRKHATSSSSTSGGGGRSSLNSIPPDQVLLLL
Liza_subviridis                   ---VAKDPEHKASAPPQDGRKRSSSCSSTGGGGGGSSVTNLPPDQVLLLL
Zoarces_viviparus                 ---ASKDLEHRT-VPSQDRRKRSSISSACVG--GKSMLTSMPPDQVLLLL
Acanthopagrus_schlegelii          ---ASKGLEHRT-APPQDRRKHIS--SSAAGGGGKSSVISMPPDQVLLLL
Halichoeres_tenuispinis           ---GSKDREQRT-VPPQGRRKHGS-----SVGGGKSPVISMPPDQVLLLL
Ictalurus_punctatus               GYSKAQSGSDVREALPQDGQ------SSSGIGGGVADVVCMSPEQVLLLL
                                     ::.   .    . .:. :              *  :  :  :.*:******

                                                             H3
Oreochromis_niloticus_niloticus   QGAEPPILSSRQKMSRPYTEVTIMTLLT[illegible]KLPGFLQ
Astatotilapia_burtoni             QGAEPPTLCSRQKMNQPYTEVTMMTLLT[illegible]KLPGFLQ
Liza_subviridis                   QGAEPPMLCSRQKLSRPYTEVTMMTLLT[illegible]KLPGFLQ
Zoarces_viviparus                 QCAEPPILCSRQKLSRPYTEVTMMTLLT[illegible]KLPGFLQ
Acanthopagrus_schlegelii          QGAEPPMLCSRQKVNRPYTEVTVMTLLT[illegible]KLPGFLQ
Halichoeres_tenuispinis           QGAEPPILCSRQKLSRPYTEVTMMTLLT[illegible]KLPGFLQ
Ictalurus_punctatus               LRAEPPTLCSRQKHSRPYSELTIMSLLT[illegible]KVPGFQD
                                   **** *.****  .:**:*:*:*:***.*:*:******:****:*** :
```

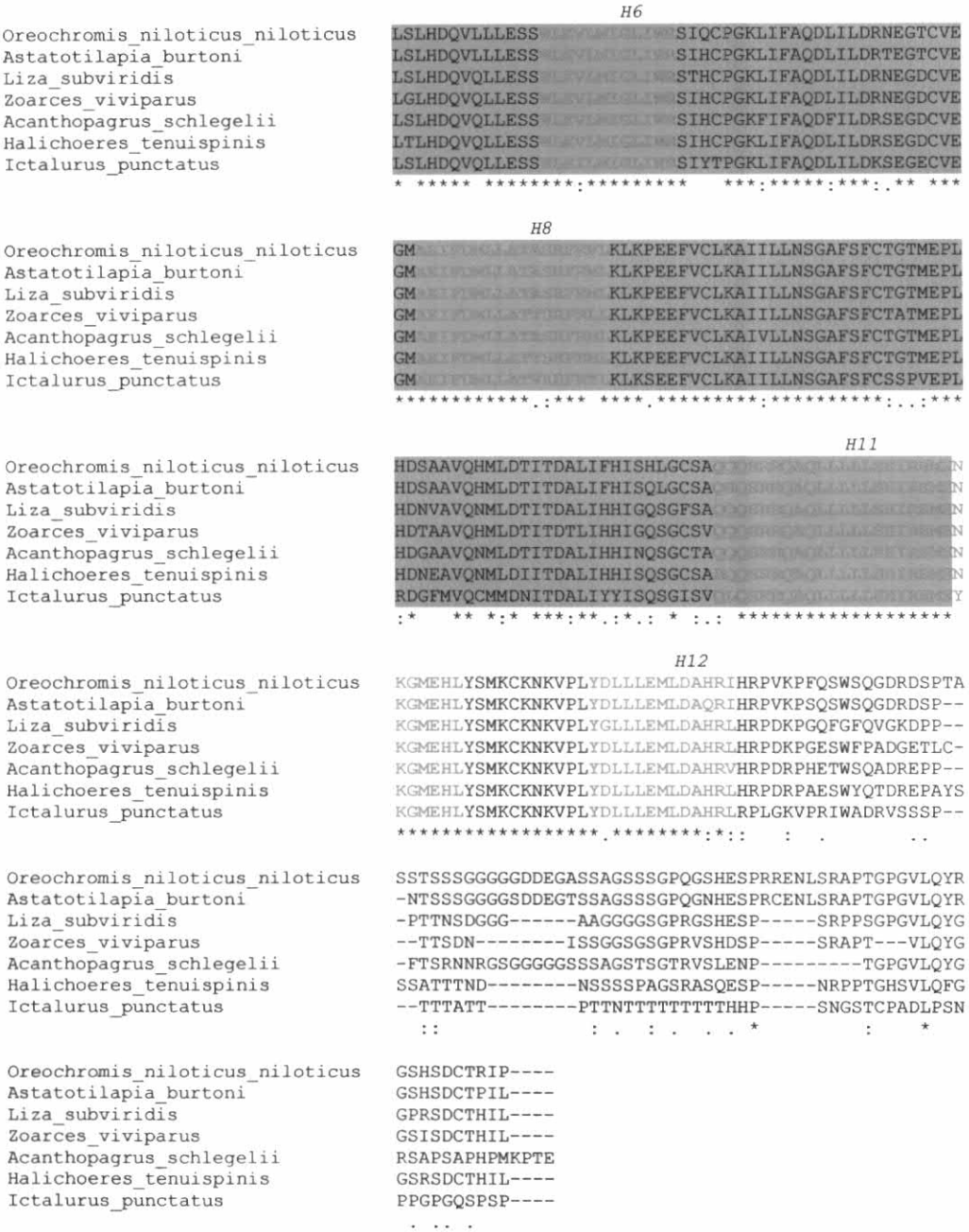


Figure 4.13 Alignment of ERα amino acid sequences of *L. subviridis* and other fish species. DNA binding domain labeled with yellow containing 2 zinc finger regions (underline), red alphabet is P box (EGCKA), and D box (PATNQ) labeled with pink. Ligand binding domain is labeled with blue. Pink alphabet is Helices H3, H6, H8, and H11 surrounding the hydrophobic cavity in tertiary structure.

Results in Fig. 4.13 showed DNA binding domain contained P box (EGCKA) directly interact with regulatory regions of DNA, and D box (PATNQ) involved in receptor dimerization, and ligand binding domain of *L. subviridis* ERα which was highly conserved with these regions of other fish ERα. Helices H3, H6, H8, and H11 surrounding the hydrophobic cavity in tertiary structure of *L. subviridis* ERα which was highly conserved with these regions of other fish ERα. Open reading frame (ORF) of *ERα* was amplified using primers designed from start and stop codon sequences (Fig. 4.11). A DNA fragment at the size of 1.9 kb was obtained from the amplification in *L. subviridis* liver (Fig. 4.14)

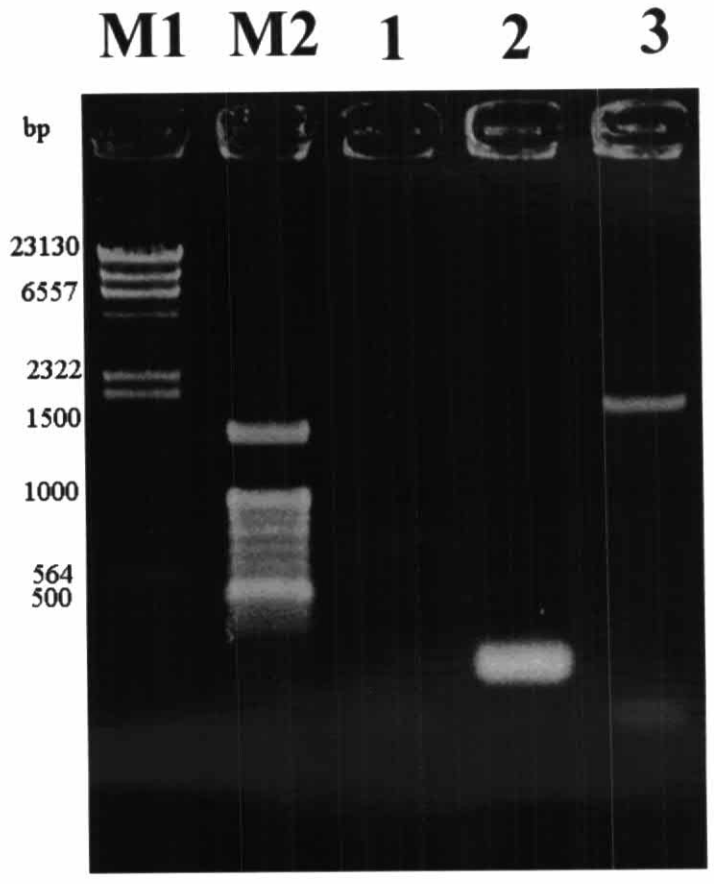


Figure 4.14 PCR product of *ERα* ORF separated in 1.2% agarose gel. Lane M1, M2, 1, 2, and 3 is λ*Hin*dIII, 100 bp DNA marker, negative control, and positive control and *ERα* ORF PCR product, respectively.

BLAST analysis of *ERα* ORF indicated that it was homologous with estrogen receptor alpha of largemouth bass (*Micropterus salmoides*) at the E-value of 4e-20. Structure of full length of ERα gene which shown position of cDNA sequence of *ERα* gene obtained from this study shown in Fig.4.15

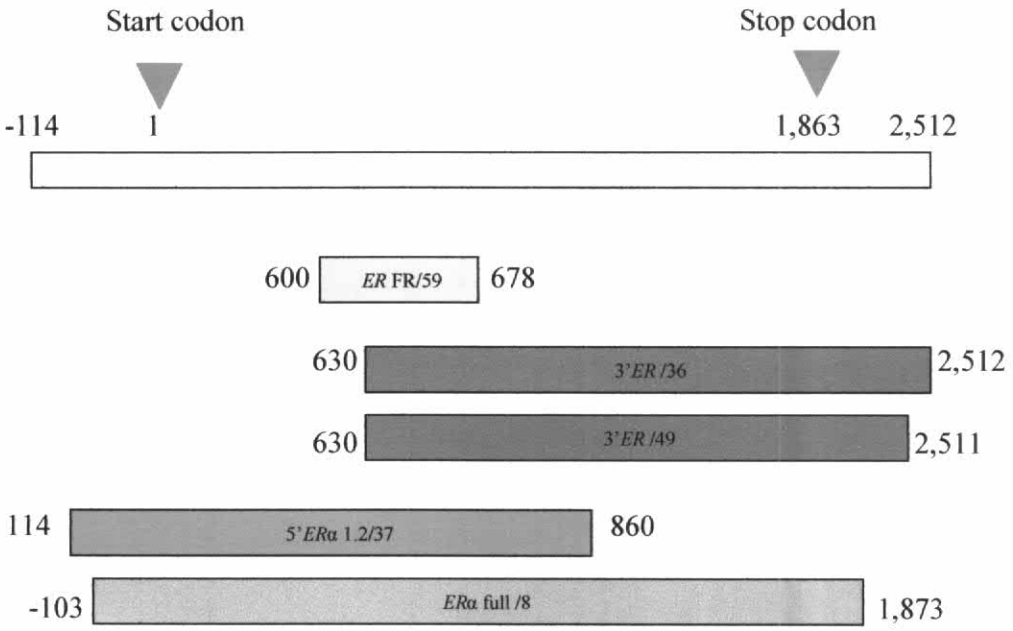


Figure 4.15 Schematic representation of the full length of *ERα* gene. Yellow box is partial coding sequence of *ERα* amplified by degenerate PCR, green box is 3'end cDNA sequence of *ERα*, blue is 5'end cDNA sequence of *ERα*, and purple box is full length ORF of *ERα*.

4.3.4 Amplification of 5' and 3'cDNA end of *ERβ* by RACE-PCR

DNA fragment at the size of 850 bp was obtained from 5'RACE PCR amplification of *ERβ* in liver (Fig.4.16). The band was cloned. Several white colonies were checked for the insert size by colony PCR. Colony PCR products of 5'RACE *ER* were shown in Fig. 4.17. Clone No.36 (insert size of 850 bp), 23 (insert size lower than 850 bp) and 39 (insert size higher than 850 bp) were cultured for plasmid extraction, sequencing and BLAST analysis indicated that the sequences of 5'*ER*/23, 36 and 39 were homologous with *ERβ* of *Micropterus salmoides* with E-value of 7e-111, 3e-145 and 5e-141, respectively (see in appendix C).

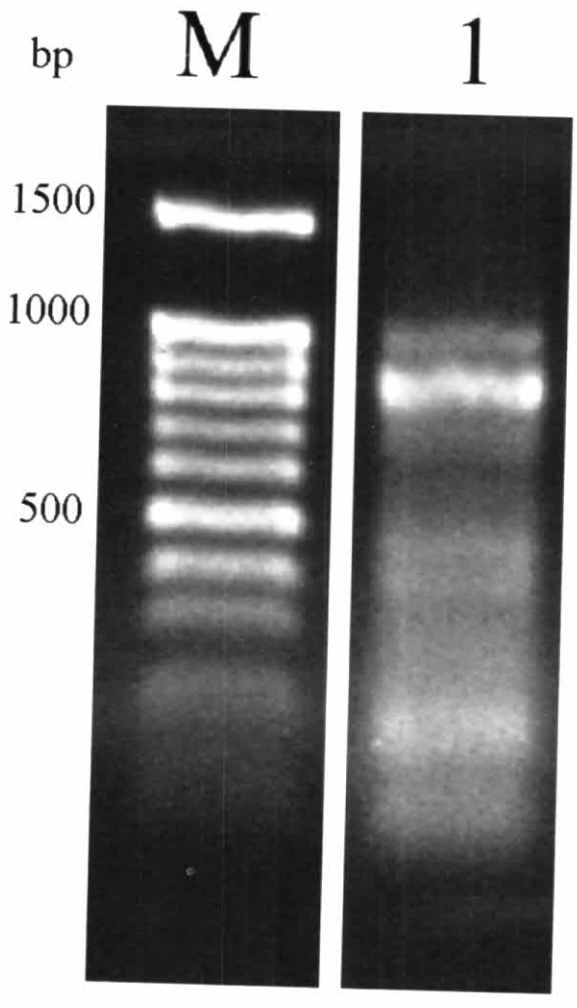


Figure 4.16 5'RACE PCR product of *ERβ* separated in 1.2% agarose gel. Lane M is 100 bp DNA marker, lane 1 is 5'RACE PCR product of *ERβ*.

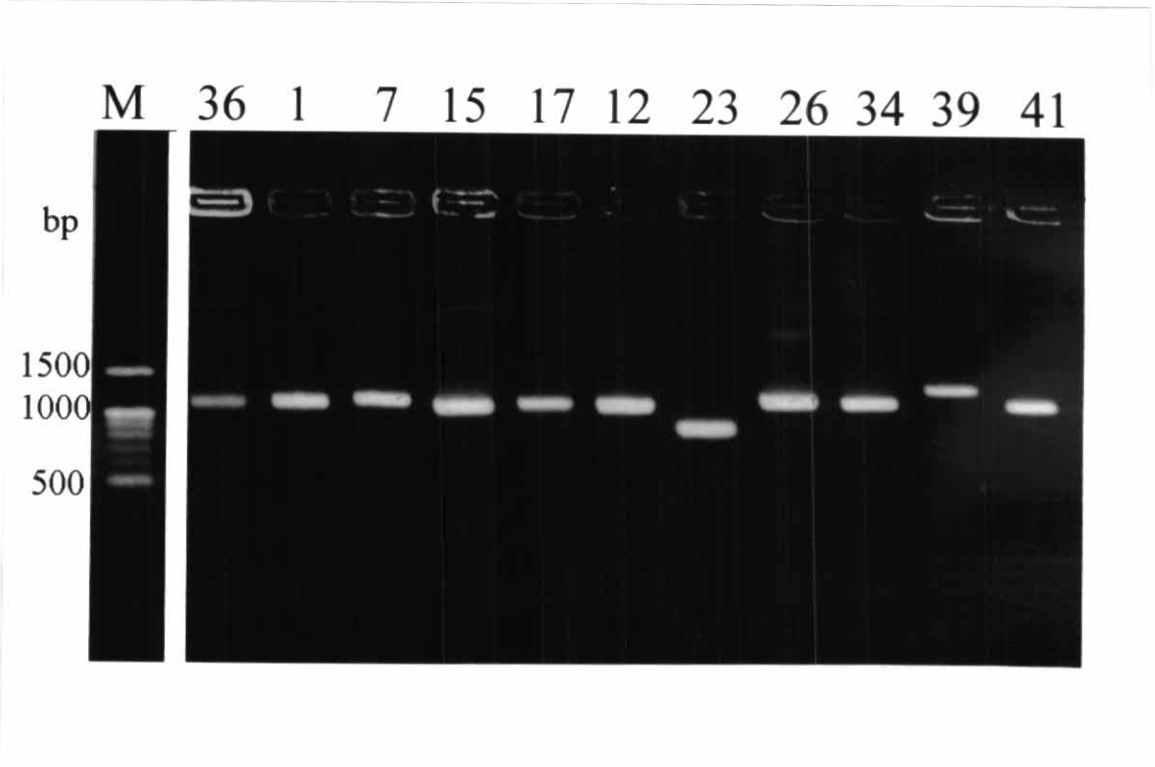


Figure 4.17 colony PCR product of 5' RACE *ER* separated in 1.2% agarose gel. Lane M is 100 bp DNA marker, number of lane represent number of clone.

Because of *ER* FR/59 sequence used for designing *ER* R-RACE primer was DNA binding domain which was very highly conserved region between *ER*α and *ER*β. 5' RACE PCR product of *ER* could amplify 2 isoforms. Therefore, 5'RACE PCR amplification of *ER* in liver was again which used BD Advantage 2 Polymerase Mix (Clontech). DNA fragment at size of 850 bp was obtained from 5'RACE PCR amplification in liver (Fig. 4.18).

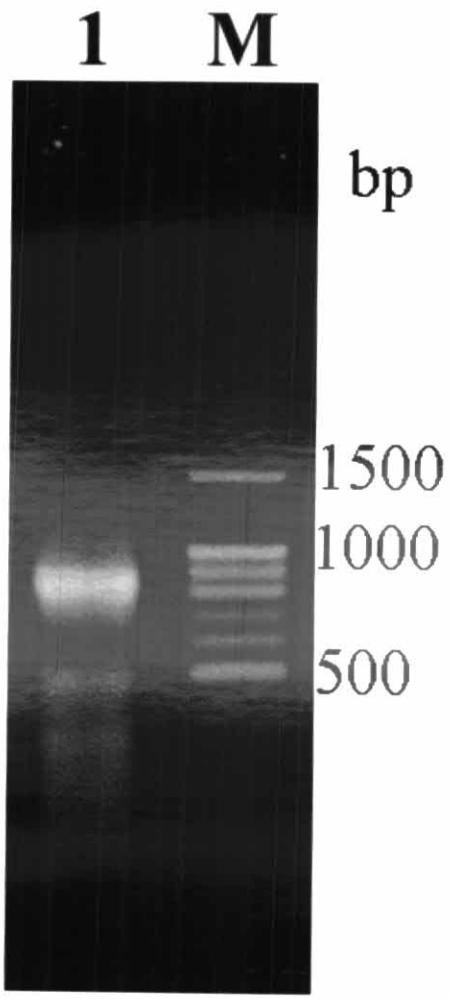


Figure 4.18 5'RACE PCR product of *ER* amplified by BD Advantage 2 Polymerase Mix (Clontech) separated in 1.2% agarose gel. Lane M is 100 bp DNA marker, lane 1 is 5'RACE PCR product of *ER* amplified by BD Advantage 2 Polymerase Mix (Clontech).

After sequencing (see in appendix B) and BLAST analysis (see in appendix C), it homologous with *ERβ* (*Micropterus salmoides*) with E-value was 1e-138. 4 types of 5'cDNA end of *ERβ* were aligned as shown in Fig. 4.19

```
5'ER-36          ------------------------------------------------------------
5'-ER-900-14     ------------------------------------------------------------
5'-ER-23         ------------------------------------------------------------
5'-ER-39         ACGCGGGGAGTGAGAGACCGAGCTCGAGAGCTCGAGGGGGGAAGAGCGGAAACGTTTGAT

5'ER-36          ------------------------------------------------------------
5'-ER-900-14     ------------------------------------------------------------
5'-ER-23         ------------------------------------------------------------
5'-ER-39         TGAGTCTCAGGAGAAGTTTGTCGGAGGAGTAAACGCCGGATGCGTCAGCTTGCGCGCTCA

5'ER-36          ------------------------------------------------------ACGCG
5'-ER-900-14     ------------------------------------------------------------
5'-ER-23         ------------------------------------------------------------
5'-ER-39         CACTTTCACTTTTACGCGCTCGTCCTTTTCCTTCTTCTTCTTCTTCTTCTTGACGTT


5'ER-36          GGG-GAAAGCGGTTGAGCCACAGAG--CCTGGACAAGAAAGTCTTCCTCCTGTTCCCGAT
5'-ER-900-14     ----------------------------ACAAGAAAGTCTTCCTCCTGTTCCCGAT
5'-ER-23         ------------------------------------------------------------
5'-ER-39         TAACGGCAGCGACGCGTTCGTGGAGGATCCGGACGAACCGAGTGAAGTCCAGCAGCTGCA

5'ER-36          GTCGTGTGAGTCTTGAAGTCGGCCTCTAGTC--GGACGTGTCCACAGAAGACGATGCAGT
5'-ER-900-14     GTCGTGTGAGTCTTGAAGTCGGCCTCTAGTC--GGACGTGTCCACAGAAGACGATGCAGT
5'-ER-23         ------------------------------------------------------------
5'-ER-39         GCAGC-TGATCCTCTGCTCCGACCTCCAGCTCAGGACGTGTCCACAGAAGACGATGCAGT

5'ER-36          AGACGAACCAGGTGCATCGTCTTATTGTCCATCACCTCCACGTTTCCTGTCCCTGACATG
5'-ER-900-14     AGACGAACCAGGTGCATCGTCTTATTGTCCATCACCTCCACGTTTCCTGTCCCTGACATG
5'-ER-23         ------------------------------------------------------------
5'-ER-39         AGACGAACCAGGTGCATCGTCTTATTGTCCATCACCTCCACGTTTCCTGTCCCTGACATG

5'ER-36          GCCTCGTCCCCTGGGCTGGATGCGGACCCGTTACCCCTGCTTCAGCTACAGGAGGTGGAC
5'-ER-900-14     GCCTCGTCTCCTGGGCTGGATGCGGACCCGTTACCCCTGCTTCAGCTACAGGAGGTGGAC
5'-ER-23         ------------------------------------------------------------
5'-ER-39         GCCTCGTCCCCTGGGCTGGATGCGGACCCGTTACCCCTGCTTCAGCTACAGGAGGTGGAC

5'ER-36          TCCAGTAAAGCCACCGAGAGGCCGAGCTCCCCAGGACTCCTGCCGGTCATGTACAGCCCT
5'-ER-900-14     TCCAGTAAAGCCACCGAGAGGCCGAGCTCCCCAGGACTCCTGCCGGTCATGTACAGCCCT
5'-ER-23         ----ACGCGGGCACCGAGAGGCCGAGCTCCCCAGGACTCCTGCCGGTCATGTACAGCCCT
5'-ER-39         TCCAGTAAAGCCACCGAGAGGCCGAGCTCCCCAGGACTCCTGCCGGTCATGTACAGCCCT
                          * **************************************************

5'ER-36          CCTGTGGGCGTGGACAGCCACACCGTCTGCATCCCTTCTCCGTACACGGACAGTAGCCAC
5'-ER-900-14     CCCGTGGGCGTGGACAGCCACACCGTCTGCATCCCTTCTCCGTACACGGACAGTAGCCAT
5'-ER-23         CCCGTGGGCGTGGACAGCCACACCGTCTGCATCCCTTCTCCGTACACGGACAGTAGCCAT
5'-ER-39         CCCGTGGGCGTGGACAGCCACACCGTCTGCATCCCTTCTCCGTACACGGACAGTAGCCAT
                 ** **********************************************************

5'ER-36          GACTACAGCCACGGACATGGACCTCTGACCTTCTACAGTCCGTCCATGCTCAGCTACACC
5'-ER-900-14     GACTACAGCCACGGACATGGACCTCTGACCTTCTACAGTCCGTCCATGCTCAGCTACACC
5'-ER-23         GACTACAGCCACGGACATGGACCTCTGACCTTCTACAGTCCGTCCATGCTCAGCTACACC
5'-ER-39         GACTACAGCCACGGACATGGACCTCTGACCTTCTACAGTCCGTCCATGCTCAGCTACACC
                 ************************************************************

5'ER-36          CGGCCGCCCATCACTGACAGCCCATCATCTCTGTGTCCCCCTCTCAGCCCCTCTGCCTTT
5'-ER-900-14     CGGCCGCCCATCACTGACAGCCCATCATCTCTGTGTCCCCCTCTCAGCCCCTCTGCCTTT
5'-ER-23         CGGCCGCCCATCACTGACAGCCCATCATCTCTGTGTCCCCCTCTCAGCCCCTCTGCCTTT
5'-ER-39         CGGCCGCCCATCACTGACAGCCCATCATCTCTGTGTCCCCCTCTCAGCCCCTCTGCCTTT
                 ************************************************************
```

```
5'ER-36          TGGCCGTCCCACAGTCACGCCAGCATGCCTTCGCTCACGCTGCACTGCACCCAGCCGCTG
5'-ER-900-14     TGGCCGTCCCACAGTCACGCCAGCATGCCTTCGCTCACGCTGCACTGCACCCAGCCGCTG
5'-ER-23         TGGCCGTCCCACAGTCACGCCAGCATGCCTTCGCTCACGCTGCACTGCACCCAGCCGCTG
5'-ER-39         TGGCCGTCCCACAGTCACGCCAGCATGCCTTCGCTCACGCTGCACTGCACCCAGCCGCTG
                 ************************************************************

5'ER-36          GTCTACAACGAACCCAACTCACACTCAGGCTGGCTGGAGCCCAAAGTCCACAGCATCAAC
5'-ER-900-14     GTCTACAACGAACCCAACTCACACTCAGGCTGGCTGGAGCCCAAAGTCCACAGCATCAAC
5'-ER-23         GTCTACAACGAACCCAACTCACACTCAGGCTGGCTGGAGCCCAAAGTCCACAGCATCAAC
5'-ER-39         GTCTACAACGAACCCAACTCACACTCAGGCTGGCTGGAGCCCAAAGTCCACAGCATCAAC
                 ************************************************************

5'ER-36          CCCAGCAGCGCCGTCATCAGCTGTAACAAGCTGCTGGGGAAGAAATCGGAGGACGGAGTG
5'-ER-900-14     CCCAGCAGCGCCGTCATCAGCTGTAACAAGCTGCTGGGGAAGAAATCGGAGGACGGAGTG
5'-ER-23         CCCAGCAGCGCCGTCATCAGCTGTAACAAGCTGCTGGGGAAGAAATCGGAGGACGGAGTG
5'-ER-39         CCCAGCAGCGCCGTCATCAGCTGTAACAAGCTGCTGGGGAAGAAATCGGAGGACGGAGTG
                 ************************************************************

5'ER-36          GAGAGGGCGAAGGAGTCCTCGTGTTCGTCGGCAGCAGGGAAAGCCGACATGCACTTCTGC
5'-ER-900-14     GAGAGGGCGAAGGAGTCCTCGTGTTCGTCGGCAGCAGGGAAAGCCGACATGCACTTCTGC
5'-ER-23         GAGAGGGCGAAGGAGTCCTCGTGTTCGTCGGCAGCAGGGAAAGCCGACATGCACTTCTGC
5'-ER-39         GAGAGGGCGAAGGAGTCCTCGTGTTCGTCGGCAGCAGGGAAAGCCGACATGCACTTCTGC
                 ************************************************************

5'ER-36          GCCGTGTGCCACGACTACGCGTCCGGGTACCACTACGGCGTCTGGTCCTGTGAGGGCTGC
5'-ER-900-14     GCCGTGTGCCACGACTACGCGTCCGGGTACCACTACGGCGTCTGGTCCTGTGAGGGCTGC
5'-ER-23         GCCGTGTGCCACGACTACGCGTCCGGGTACCACTACGGCGTCTGGTCCTGTGAGGGCTGC
5'-ER-39         GCCGTGTGCCACGACTACGCGTCCGGGTACCACTACGGCGTCTGGTCCTGTGAGGGCTGC
                 ************************************************************

5'ER-36          AAGGCCTTCTTCAAGAGGAGCATCCAGGGTCACAATG
5'-ER-900-14     AAGGCCTTCTTCAAGAGGAGCATCCAGGGTCACAATG
5'-ER-23         AAGGCCTTCTTCAAGAGGAGCATCCAGGGTCACAATG
5'-ER-39         AAGGCCTTCTTCAAGAGGAGCATCCAGGGTCACAATG
                 ************************************
```

Figure 4.19 Alignments of 4 types of 5'cDNA end of *ER*β. Red is start codon

Results in Fig. 4.19 showed almost identical in coding sequences but differences in 5'UTR sequences of 5'cDNA ends between 4 types of *ER*β. Further amplification of was conducted by RACE PCR using *ER*β F RACE primer designed from the obtained 5'cDNA sequence of *ER*β. DNA fragments at the size of 1.3 kb and 600 bp were obtained (Fig. 4.20). After sequencing (see in appendix B) and BLAST analysis (see in appendix C), these PCR products were homologous with *ER*β of Atlantic croaker (*Micropogonias undulatus*) and largemouth bass (*Micropterus salmoides*) with E-value of 0.0 and 3e-32, respectively.

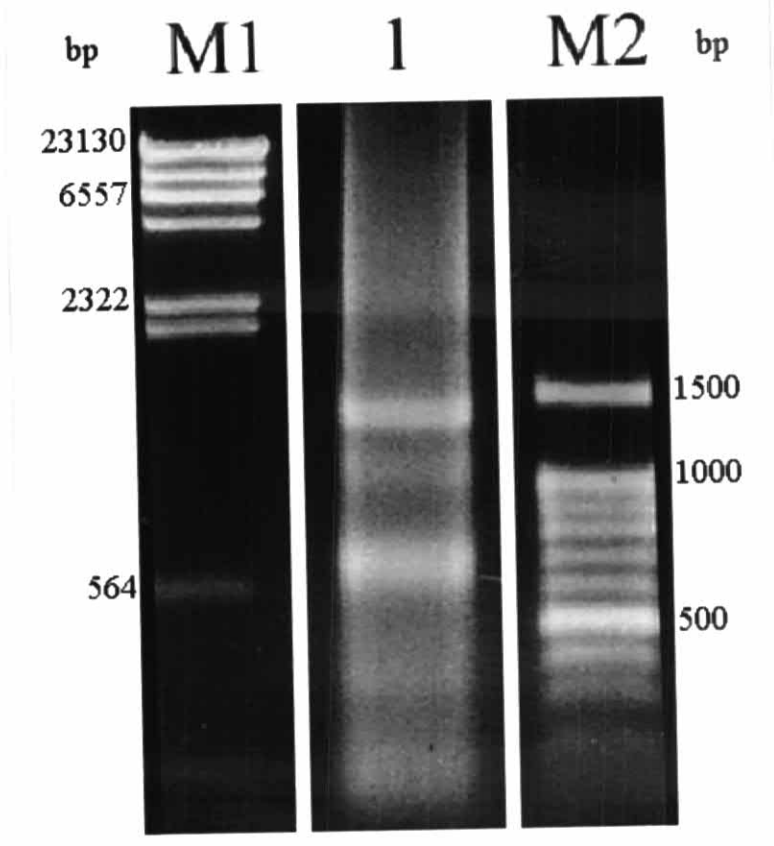


Figure 4.20 3' RACE PCR product of *ERβ* separated in 1.2 % agarose gel. Lane M1, M2 and 1 are λ*Hin*dIII, 100 bp DNA marker and 3' RACE PCR product of *ERβ*, respectively.

ERβ F1-RACE primer designed from 3'*ERβ* 1.3/33 sequence for used in 3' RACE PCR amplification. DNA fragments at size 600 bp and large than 1.5 kb were obtained from 3' RACE PCR amplification (Fig. 4.21).

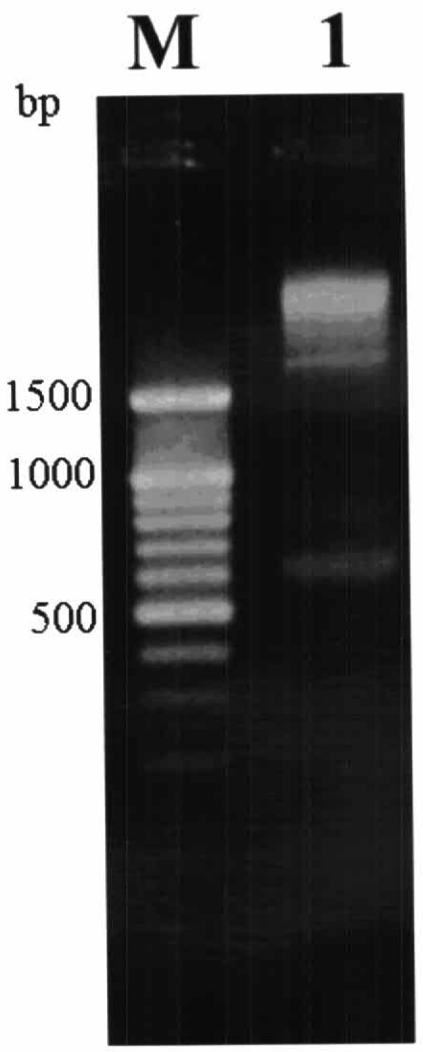


Figure 4.21 3' RACE PCR product of *ERβ* separated in 1.2 % agarose gel. Lane M and 1 are 100 bp DNA marker and 3'RACE PCR product of *ERβ*, respectively.

A DNA fragment at size 600 bp was cloned. After sequencing (see in appendix B) and BLAST analysis (see in appendix C), it was homologous with estrogen receptor beta b of *Astatotilapia burtoni* with E-value of 1e-26. 3' cDNA end sequences of *ERβ* were aligned as shown in Fig. 4.22 and 4.23

```
3'ERbeta600-38      TGACTACATTTGCCCCGCCACAAATCAGTGCACTATCGACAAGAATCGGCGTAAAAGCTG
3'ERbeta1.3-33      TGGCTACATTTGCCCCGCCACAAATCAGTACACTATCGGCAAGAATCGGCGTAAAAGCTG
                    ** ************************** ******** **********************

3'ERbeta600-38      CCAAGCCTGCCGCCTGCGGAAATGCTACGAAGTGGGCGTGATGAAGGGCCTAGCTGCCAA
3'ERbeta1.3-33      CCAAGCCTGCCGCCTGCGGAAATGCTACGAAGTGGGCATGATGAAGTG--TGGCG--TGA
                    ************************************ ******** *  * **     *

3'ERbeta600-38      ACCTTAATTGTCCTGTCCAGGGAGGCACTAGCAAACTGATTGTTGTCCTTGGGGTTGATG
3'ERbeta1.3-33      GGCGTGAACGCTGCAGCTATCGAGGAGCTCG-ACACCG-TCGTGGT----GGAGTCCAGT
                      * * *  *       * *  ****  ** * * ** * * ** **    ** **  *

3'ERbeta600-38      GACACACCTTCACACACCTG---TCTCTGAGTATGAACACATGGTTTGAGACAAAAAAAA
3'ERbeta1.3-33      CTCGGGACGCCACGGGCCGGGGCTTGGTGAAGGTCGGCCCCGGTTCTCGGGCCCAGCGGC
                      *    *  ***   ** *   *   ***   *   * *  * * *  * *  *

3'ERbeta600-38      ATATCAAAGAAAGGCAGGAAAAAATATAAAATAAAATC--TGCTTAATGTTCAAAAAACA
3'ERbeta1.3-33      ATCTCGACCTGGGACCGCCCCTG-TCCCCGCTGGCCTCCCTGCCCCAGGCCAACCACCTG
                    ** ** *     * * *       *      *    **  ***   * *   *  *

3'ERbeta600-38      CATT--TCACCATTGAGTTGATACAAACT-GTCTTCTGTCCTACGCAA-TAAACACACCA
3'ERbeta1.3-33      CACCACTCAGCCATGAGCCCGGAGGAGTTCATCTCCCGCATCATGGAGGCCGAGCCTCCA
                    **    *** *  **** *  *  *  *** * *    * * *    *  * ***

3'ERbeta600-38      CGGAACAATCCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA------------
3'ERbeta1.3-33      GAGATCTACCTGATGGAGGACCTGAAGAAGCCGTTCACCGAGGCCAGCATGATGATGTCC
                      ** * * *  *   *  *    ** **       *   *    *

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      CTCACCAACCTGGCAGACAAGGAGCTGGTGCTCATGATCAGCTGGGCCAAAAAGATCCCT

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      GGCTTCGTAGAGCTGAGTCTAGCAGATCAGATCCACCTGCTGAAGTGCTGCTGGCTGGAG

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      ATCCTCATGCTGGGTCTGATGTGGAGGTCTGTGGATCATCCTGGAAAACTCATCTTCTCT

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      CCAGACTTCAAACTCAACAGGGAGGAGGGCCAGTGTGTGGAGGGCATCATGGAGATCTTT

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      GACATGCTGCTGGCGGCCACGTCTCGCTTCAGAGAGCTGATGCTGCAGAGGGAGGAGTAC

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      GTCTGCCTGAAGGCCATGATCCTCCTCAACTCCAGTGAGTTTAACACCTCAACTCCAGTG

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      AGTTTAACACGGACGATCAGCTGATGCTACTTCACTGAGAAACACTGACAACACGACGCA

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      AACGTTCTGATTTAAATGAGCCACCACAGGCCACGCCCACGGCCAGAAGAGGGCGTGGCC

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      TGTGGATGTGAGGACAAAAAAACAAGGCCCCGCCCACATGTGAAATTATTCTCTCATTGT

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      CCATGTCTCTGTGGAGGACAGTTGTTCTCTAGTGGACGTCCCAGGCTCTGGTTGCCATGG
```

```
3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      TAACCCCCATATAAACATGTCTAACTCTTCTTCTGGGTTCAGCCTCCAAGCTGCTGTTGT

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      CACGGTGGGATGGACGGTGTCTCCACCTTAAATAAACCTATAGTCCACCTTAAGTGTCAA


3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      TCAAATGGATCGGTGGCTCCCATTGGTCCACTTCAGCTTGGTACCGCCCACTCTGTGTCT

3'ERbeta600-38      ------------------------------------------------------------
3'ERbeta1.3-33      GGTCTCTTCCCTTTAAAAGTCACTTTTAAATTGGTGAAAAAAGTTAAAGAAATAAGGGAA

3'ERbeta600-38      ---------------------------------
3'ERbeta1.3-33      GGATAATAAATAAAAAAAAAAAAAAAAAAAAAA
```

Figure 4.22 Alignment of 3'*ER*β 1.3-33 and 600/38. Blue and red alphabet shown nucleotides which different, green, blue and yellow label is stop codon, polyadenylation signal and polyA tail, respectively.

```
3'ER-beta-600-38      ------------------------------------------------------------
3'ER-beta-600-1       GCGGCCACCTCTCGCTTCAGAGAGCTGAAGCTGCAGAGGGAGGAGTACGTCTGCCTGAAG

3'ER-beta-600-38      ------------------------------------------------------------
3'ER-beta-600-1       GCCATGATCCTCCTCAACTCCAGTGAGTTTAACACGGACGATCAGCTGATGCTGCTTCAC

3'ER-beta-600-38      ---------------CTTCAAGAGGAGCATCCAGGGTCACAATGACTAC-ATTTGCCCCG
3'ER-beta-600-1       TGAGAAACACTGACAACACGACGCAAACGTTCTGATTTAAATGAGCCACCACAGGCCACG
                                   * *    * * * * *  * * *    * ** *   *** **

3'ER-beta-600-38      CCA-----CAAATCAGTGCACTATCGACAAGAATCGGCGTAAAAGCTGCCAAGCCTGCC-
3'ER-beta-600-1       CCCACGGCCAGAAGAGGGCGTGGTCTGTGGATGTGAGGACAAAAAAACAAGGCCCCGCCC
                      **      ** *  ** **    **        *  *   ****          ** ***

3'ER-beta-600-38      GCCTGCGGAAATGCTA-CGAAGTGGGCGTG-----ATGAAGGGC---------CTAGCTG
3'ER-beta-600-1       ACATGTGAAATTATTCTCTCATTGTCCATGTCTCTGTGGAGGACAGTTGTTCTCTAGTGG
                       * ** * ** *  *  *  * **  * **      ** *** *         ****  *

3'ER-beta-600-38      -----CCAAACCTTAATTGTCCTGTCCAGGGAGGCACTAGCAAA-CTGATTGTTGTCCTT
3'ER-beta-600-1       ACGTCCCAGGCTCTGGTTGCCATGGTAACCCCCATATAAACATGTCTAACTCTTCTTCTG
                           ***  *  *  *** * **   *       *  * **   ** * * ** * **

3'ER-beta-600-38      GG-----------GGTTGATGGACACACCTTCACACAC-CTGTCTCTGAGTATGAACACA
3'ER-beta-600-1       GGTTCAGCCTCCAAGCTGCTGTCGTCACGGTGGGATGGACGGTGTCTCCACCTTAA-ATA
                      **            * ** **    ***  *   *    * ** ***     * ** * *

3'ER-beta-600-38      TGGTTTGAGACAAAAAAAAATATCAAAGAAAGGCAGGAAAAAAT---ATAAAATAAAATC
3'ER-beta-600-1       AACCTATAGTCCACCTTAAGTGTCAAACAAATGGATCGGTGGCTCCCATTGGTCCAGTTC
                          *  ** * *    ** * ***** *** * *        *   **      *  **

3'ER-beta-600-38      TGCTTAATGTT-CAAAAAACACATTTCACCATTGAGTTGATACAAACTGTCTTCTGTCCT
3'ER-beta-600-1       AGCTTGGTACCGCCCACTCTGTGTCTGGTCTCTTCCCTT-TAAAAGTCACTTTTAAATTG
                       ****  *    *  *       * *   *  *    *  ** **      **

3'ER-beta-600-38      ACGCAA------TAAACACACCACGGAACAATCCTAAAAAAAAAAAAAAAAAAAAAAAAA
3'ER-beta-600-1       ATGGAAAAAAGTTAAAGAAATAAGGGAAGGATAATAAATAAAAAAAAAAAAAAAAAAAAA
                      * * **      **** * *  * ****  **  **** **********************

3'ER-beta-600-38      AAAAAAAAAA
3'ER-beta-600-1       AAAAAAAAAA
                      **********
```

Figure 4.23 Alignment of 3'*ER*β 600/1 and 600/38

Results in Fig. 4.22 and 4.23 showed different in nucleotide sequences of 3'cDNA end of *ER*β, but we not determined additionally sequence (see in Fig. 4.23).

4.3.5 Full length cDNA sequence determination of *ER*β

5'*ER*/36 and 3'*ER*β 1.3/33 cDNA sequence were connected *in silico* and obtained full length mRNA of *ER*β at size 2,098 bp which contained 5'UTR, ORF, and 3'UTR at size 177, 1,431, and 490 bp, respectively. Domains and motifs found in the sequence showed in Fig. 4.24 and table 4.2. *ER*β ORF encoded ERβ contained 476 amino acid residues with the molecular weight at 52.36 kDa which contained A/B domain, DNA binding domain, D domain and ligand binding domain but lack F domain. DNA binding domain was found at 178-253 with score as 17.312. Two zinc finger regions were found at 181-201 and 217-236. P box was

found at 199-204. D box was found at 218-222. Ligand binding domain was found at 352-476 with E-value as 6.80e-15 as shown in Fig. 4.24.

Table 4.2 Domains, motifs, and consensus patterns of ERβ

Domain	Site	Consensus pattern
DNA-binding domain	178-253	C - x(2) - C - x(1,2) - [DENAVSPHKQT] - x(5,6) - [HNY] - [FY] - x(4) - C - x(2) - C - x(2) - F(2) - x – R The 4 C's are zinc ligands (http://br.expasy.org/prosite)
P box	199-203	EGCKA (Attwood, Kroll, and Denslow, 2004)
D box	218-222	PATNQ (Attwood, Kroll, and Denslow, 2004)
zinc finger region (NR C4-type)	181-201, 217-236	C - x(2) - C - x(1,2) - [DENAVSPHKQT] - x(5,6) - [HNY] - [FY] - x(4) - C - x(2) - C - x(2) - F(2) - x - R (http://br.expasy.org/prosite)
ligand binding domain	352-476	Contained helices H3, H6, H8, and H11 surrounding the hydrophobic cavity in tertiary structure (Menuet et al, 2002)

```
acgcgggggaaagcggttgagccacagagcctggacaagaaagtcttcctcctgttcccg
atgtcgtgtgagtcttgaagtcggcctctagtcggacgtgtccacagaagacgatgcagt
agacgaaccaggtgcatcgtcttattgtccatcacctccacgtttcctgtccctgacatg
                                                          M
gcctcgtcccctgggctggatgcggacccgttacccctgcttcagctacaggaggtggac
 A  S  S  P  G  L  D  A  D  P  L  P  L  L  Q  L  Q  E  V  D
tccagtaaagccaccgagaggccgagctccccaggactcctgccggtcatgtacagccct
 S  S  K  A  T  E  R  P  S  S  P  G  L  L  P  V  M  Y  S  P
cctgtgggcgtggacagccacaccgtctgcatcccttctccgtacacggacagtagccac
 P  V  G  V  D  S  H  T  V  C  I  P  S  P  Y  T  D  S  S  H
gactacagccacggacatggacctctgaccttctacagtccgtccatgctcagctacacc
 D  Y  S  H  G  H  G  P  L  T  F  Y  S  P  S  M  L  S  Y  T
cggccgcccatcactgacagcccatcatctctgtgtccccctctcagcccctctgccttt
 R  P  P  I  T  D  S  P  S  S  L  C  P  P  L  S  P  S  A  F
tggccgtcccacagtcacgccagcatgccttcgctcacgctgcactgcacccagccgctg
 W  P  S  H  S  H  A  S  M  P  S  L  T  L  H  C  T  Q  P  L
gtctacaacgaacccaactcacactcaggctggctggagcccaaagtccacagcatcaac
 V  Y  N  E  P  N  S  H  S  G  W  L  E  P  K  V  H  S  I  N
cccagcagcgccgtcatcagctgtaacaagctgctggggaagaaatcggaggacggagtg
 P  S  S  A  V  I  S  C  N  K  L  L  G  K  K  S  E  D  G  V
gagagggcgaaggagtcctcgtgttcgtcggcagcagggaaagccgacatgcacttctgc
 E  R  A  K  E  S  S  C  S  S  A  A  G  K  A  D  M  H  F  C
gccgtgtgccacgactacgcgtccgggtaccactacggcgtctggtcctgtgagggctgc
 A  V  C  H  D  Y  A  S  G  Y  H  Y  G  V  W  S  C  E  G  C
aaggccttcttcaagaggagcatccagggtcacaatggctacatttgccccgccacaaat
 K  A  F  F  K  R  S  I  Q  G  H  N  G  Y  I  C  P  A  T  N
cagtacactatcggcaagaatcggcgtaaaagctgccaagcctgccgcctgcggaaatgc
 Q  Y  T  I  G  K  N  R  R  K  S  C  Q  A  C  R  L  R  K  C
tacgaagtgggcatgatgaagtgtggcgtgaggcgtgaacgctgcagctatcgaggagct
 Y  E  V  G  M  M  K  C  G  V  R  R  E  R  C  S  Y  R  G  A
cgacaccgtcgtggtggagtccagtctcgggacgccacgggccggggcttggtgaaggtc
 R  H  R  R  G  G  V  Q  S  R  D  A  T  G  R  G  L  V  K  V
ggccccggttctcgggcccagcggcatctcgacctgggaccgcccctgtccccgctggcc
 G  P  G  S  R  A  Q  R  H  L  D  L  G  P  P  L  S  P  L  A
tccctgccccaggccaaccacctgcaccactcagccatgagcccggaggagttcatctcc
 S  L  P  Q  A  N  H  L  H  H  S  A  M  S  P  E  E  F  I  S
cgcatcatggaggccgagcctccagagatctacctgatggaggacctgaagaagccgttc
 R  I  M  E  A  E  P  P  E  I  Y  L  M  E  D  L  K  K  P  F
accgaggccagcatgatgatgtccctcaccaacctggcagacaaggagctggtgctcatg
 T  E  A  S  M  M  M  S  L  T  N  L  A  D  K  E  L  V  L  M
atcagctgggccaaaaagatccctggcttcgtagagctgagtctagcagatcagatccac
 I  S  W  A  K  K  I  P  G  F  V  E  L  S  L  A  D  Q  I  H
```

```
ctgctgaagtgctgctggctggagatcctcatgctgggtctgatgtggaggtctgtggat
 L  L  K  C  C  W  L  E  I  L  M  L  G  L  M  W  R  S  V  D
catcctggaaaactcatcttctctccagacttcaaactcaacagggaggagggccagtgt
 H  P  G  K  L  I  F  S  P  D  F  K  L  N  R  E  E  G  Q  C
gtggagggcatcatggagatctttgacatgctgctggcggccacgtctcgcttcagagag
 V  E  G  I  M  E  I  F  D  M  L  L  A  A  T  S  R  F  R  E
ctgatgctgcagagggaggagtacgtctgcctgaaggccatgatcctcctcaactccagt
 L  M  L  Q  R  E  E  Y  V  C  L  K  A  M  I  L  L  N  S  S
gagtttaacacctcaactccagtgagtttaacacggacgatcagctgatgctacttcact
 E  F  N  T  S  T  P  V  S  L  T  R  T  I  S
gagaaacactgacaacacgacgcaaacgttctgatttaaatgagccaccacaggccacgc
ccacggccagaagagggcgtggcctgtggatgtgaggacaaaaaaacaaggccccgccca
catgtgaaattattctctcattgtccatgtctctgtggaggacagttgttctctagtgga
cgtcccaggctctggttgccatggtaacccccatataaacatgtctaactcttcttctgg
gttcagcctccaagctgctgttgtcacggtgggatggacggtgtctccaccttaaataaa
cctatagtccaccttaagtgtcaatcaaatggatcggtggctcccattggtccacttcag
cttggtaccgcccactctgtgtctggtctcttcccttaaaagtcacttttaaattggtg
aaaaaagttaaagaaataagggaaggataataaataaaaaaaaaaaaaaaaaaaaaaaa
```

Figure 4.24 Full length mRNA and deduced sequence of *ER*β. Pink label, yellow label, blue alphabet, red alphabet, blue label, and green label is start codon, DNA binding domain, P box, D box, ligand binding domain, and stop codon, respectively.

ERβ sequence of *L. subviridis* was aligned with ERβ of other fish species. The result was shown in Fig. 4.25

```
AAG16711            MASSPGLDPHPLPMLQLQEVGSSKVSERPRSPGLLPAVYSPPLGMDSHTVCIPSPYTDSS
AAO39210            MASSPGLDADPLPLLQLQEVDSSKASQRPSSPGLLPAVYSPPLGMDSHTVCIPSPYTDSS
AAP72179            MASSPGLDTDPLPLLRLQEVDSSKASERPSSPGLLPAVYSPRQGMDSHTVYIPSPYTDNN
Liza_subviridis     MASSPGLDADPLPLLQLQEVDSSKATERPSSPGLLPVMYSPPVGVDSHTVCIPSPYTDSS
ABC68616            MASSPGLNAEPLPMLQLQEAYSSKPSERPTSPGLLPTMYSPPLGIDSHTVCIPSPYTDSS
                    *******:..***:*:***. *** ::** ******.:***  *:***** *******..

AAG16711            HEYNHSHGPLTFYSPSVLSYSRPPITNSPSSLCPSLSPSAFWPSHNHPTMPSLTLHCPES
AAO39210            HEYNHSHGPLTFYSPSVLSYARPPITDSPSSLCPPLSPSAFWPSHSHPNMPSLTLRCPQP
AAP72179            QEYNHGSGSVSFYSPSVLSYARPSATDSPSSLCGPLSPSAFWPPHSQPNLPSLTLRCPQP
Liza_subviridis     HDYSHGHGPLTFYSPSMLSYTRPPITDSPSSLCPPLSPSAFWPSHSHASMPSLTLHCTQP
ABC68616            HDYNHGHGPLTFYGPSVLSYSRPPITDSPTSLCPPRSPSAFWSSHSHHNVPSLTLHCTQP
                    ::*.*. *.::**.**:***:**. *:**:*** . ******..*.: .:*****:*.:.

AAG16711            IVYNEPSPHAPWLESKAHSINASSSSIIGCNKSLVKRSEEGVEDMNSSLCSSAVGKADMH
AAO39210            LVYNEPSPHAHWPEPKAHSINPSSS-ILGCNKPLGKRLEEGVEGVNSSLCSSAVGKADMH
AAP72179            LGYNESGLHAPWLESKPHNISSSSS-IIGCNKPLGKRSEEGVNGVNPSLCSSVVGKADMH
Liza_subviridis     LVYNEPNSHSGWLEPKVHSINPSSA-VISCNKLLGKKSEDGVERAKESSCSSAAGKADMH
ABC68616            LVYSEPGPHPAWLNPKAHSINPNSS-VISCNRLLGKKPDEGVEGVKSSSCSSAAGKADMH
                    : *.*.. *. * :.* *.*...*: ::.**: * *: ::**:  : * ***..******

AAG16711            FCAVCHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLR
AAO39210            FCAVCHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLR
AAP72179            FCAVCHDYASGYHYGVWSCEGCKAFFKRSIQGHNDYICPATNQCTIDKNRRKSCQACRLR
Liza_subviridis     FCAVCHDYASGYHYGVWSCEGCKAFFKRSIQGHNGYICPATNQYTIGKNRRKSCQACRLR
ABC68616            FCDVCHDYASGYHYGVWSCEGCKAFFKRSIQRHNDYICPATNQCTIDKNRRKSCQACRLR
                    ** ************************** **.******** **.************

AAG16711            KCYEVGMMKCGVRRERCSYRGARHRRGGLQPRDPTGRGLVRVGLGSRAQRHLHLEAPLTP
AAO39210            KCYEVGMMKCGVRRERCSYRGARHRRGGLQPRDPTGRGLVRMGLGSRAQRHLHLEAPLAP
AAP72179            KCYEVGMMKCGVRRERCSYRGTRHRRGGLQPRDPTGRGLVRVGLGSRAQRHLHLEGPLTP
Liza_subviridis     KCYEVGMMKCGVRRERCSYRGARHRRGGVQSRDATGRGLVKVGPGSRAQRHLDLGPPLSP
ABC68616            KCYEVGMMKCGVRRERCSYRGARHRRGGLQPR-----GLVRIGIGSRAQRLPPIELPFSS
                    **********************:******:*.*     ***::* ******   :  *:::.
```

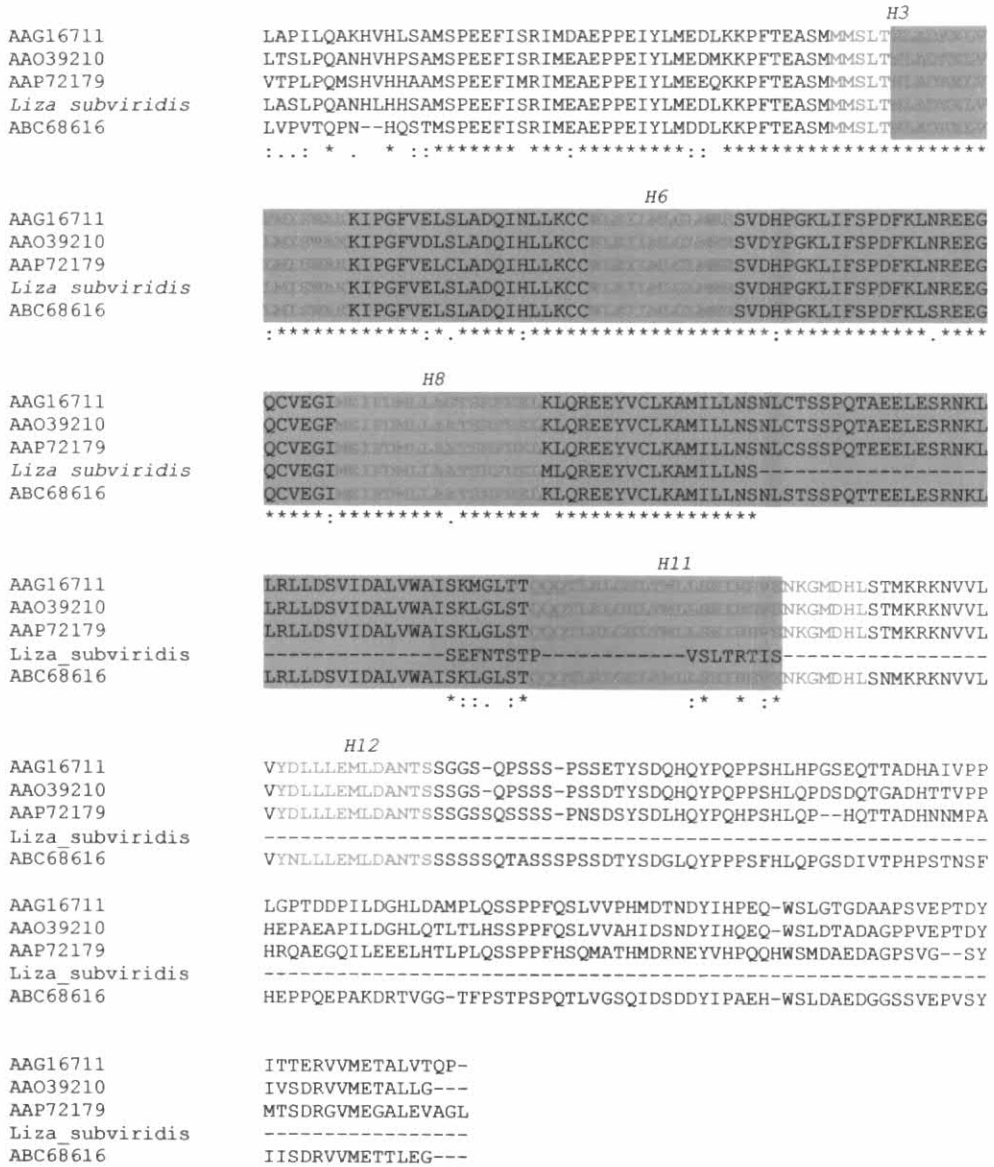


Figure 4.25 Alignment of ERβ amino acid sequence of *L. subviridis* and other species of fish. DNA binding domain is labeled with yellow. Two zinc finger regions indicate with underline, red alphabet is P box, D box is labeled with pink, ligand binding domain is labeled with blue, and helices H3, H6, H8, H11, and H12 surrounding the hydrophobic cavity in tertiary structure is pink alphabet.

Results in Fig. 4.25 showed highly conserved of amino acid sequences in DNA binding domain and ligand binding domain of fish ERβ. *L. subviridis* ERβ lack some region of ligand binding domain, helices H11 and H12 surrounding the hydrophobic cavity in tertiary structure and F domain. Structure of full length of *ERβ* gene which shown position of cDNA sequence of *ERβ* gene obtained from this study shown in Fig.4.26

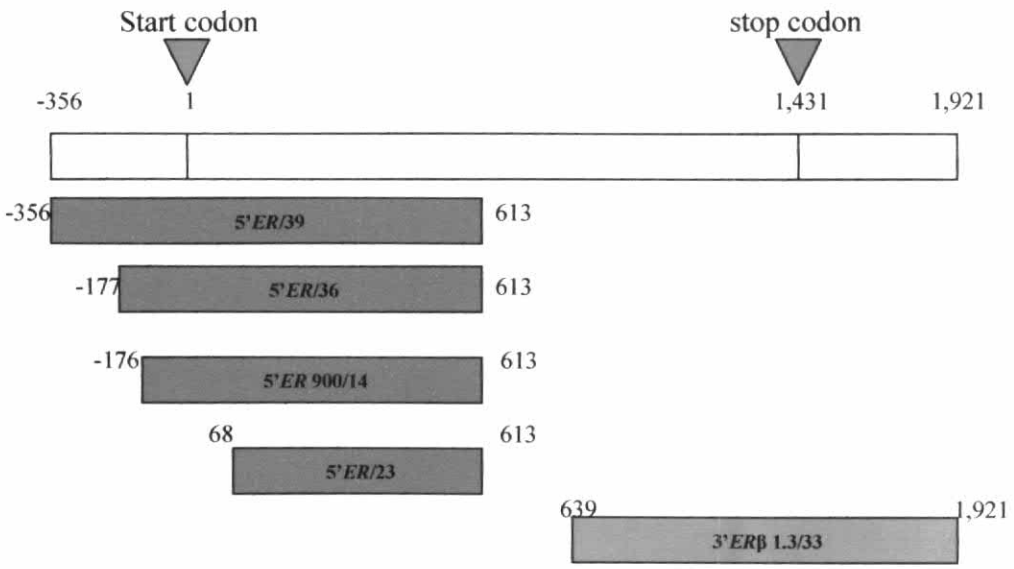


Figure 4.26 Schematic representation of the full length of *ERβ* gene. Green box is 5'end cDNA sequence of *ERβ*, blue box is 3'end cDNA sequence of *ERβ*.

4.4 Cloning and characterization of choriogenin (*chg*) genes

4.4.1 choriogenin L (*chg*-L)

Partial cDNA sequence of *chg*-L gene was amplified from 1st strand cDNA templates of the fish liver. PCR products were determined by electrophoresis on 1.2% agarose (w/v) (Fig.4.27). DNA fragments at the sizes of 900 and 700 bp were obtained from PCR amplification. PCR product at the size of 900 bp which was close to the expected size of the product (921 bp) was eluted, reamplified, and determined by electrophoresis on 1.2% agarose (w/v). After sequencing (see in appendix B) and BLAST analysis (see in appendix C), the result indicated that this PCR product was homologous with chorion protein of gilthead seabream (*Sparus aurata*) with E-value of 2e-108. Further amplification was conducted by 5' and

3'RACE PCR using F RACE primer and GW R primer designed from the obtained partial sequence of *chg*-L. Two DNA fragments at the same size of 450 bp were obtained from 5' (Fig. 4.28 A) and 3' (Fig. 4.28 B) RACE PCR amplification. After sequencing of 2 different sizes of insert of 5' RACE PCR products of *chg*-L (5'L 500/12 and 18) and 3'RACE PCR product of *chg*-L (3'L 500/19 and 37) (see in appendix B) and BLAST analysis, these sequence were homologous with chorion protein of gilthead seabream (*Sparus aurata*), chorion protein of gilthead seabream (*Sparus aurata*), choriogenin L of Javanese rice fish (*Oryzias javanicus*), and choriogenin L of Javanese rice fish (*Oryzias javanicus*) with E-value as 3e-14, 4e-22, 1e-04, and 1e-04 (see details in appendix C). Two clones of 5' and 3' cDNA ends of *chg*-L were aligned as shown in Fig. 4.29 and 4.30, respectively.

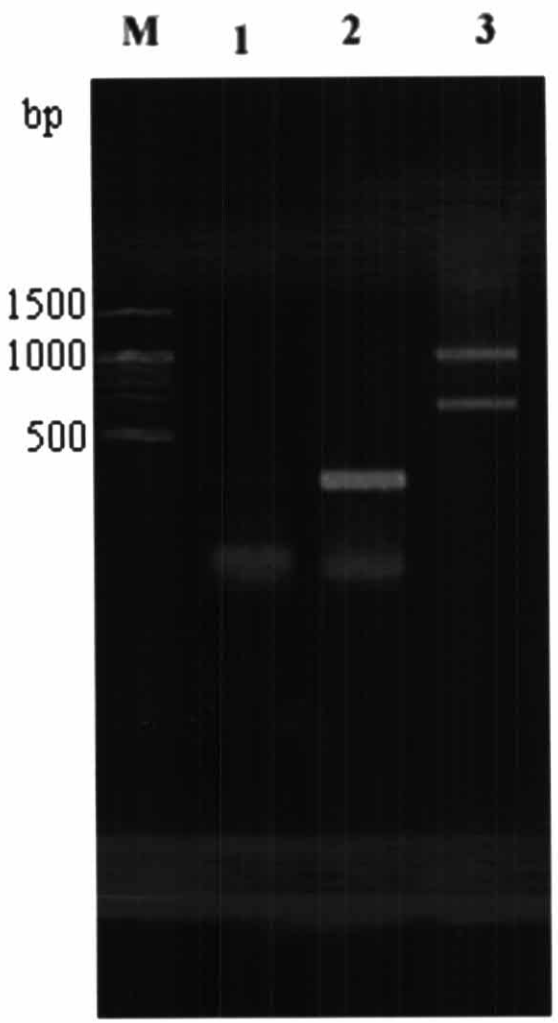


Figure 4.27 *chg*-L PCR products separated in 1.2 % agarose gel. Lane M, 1, 2 and 3 are 100 bp DNA marker, negative control, positive control and *chg*-L PCR product, respectively.

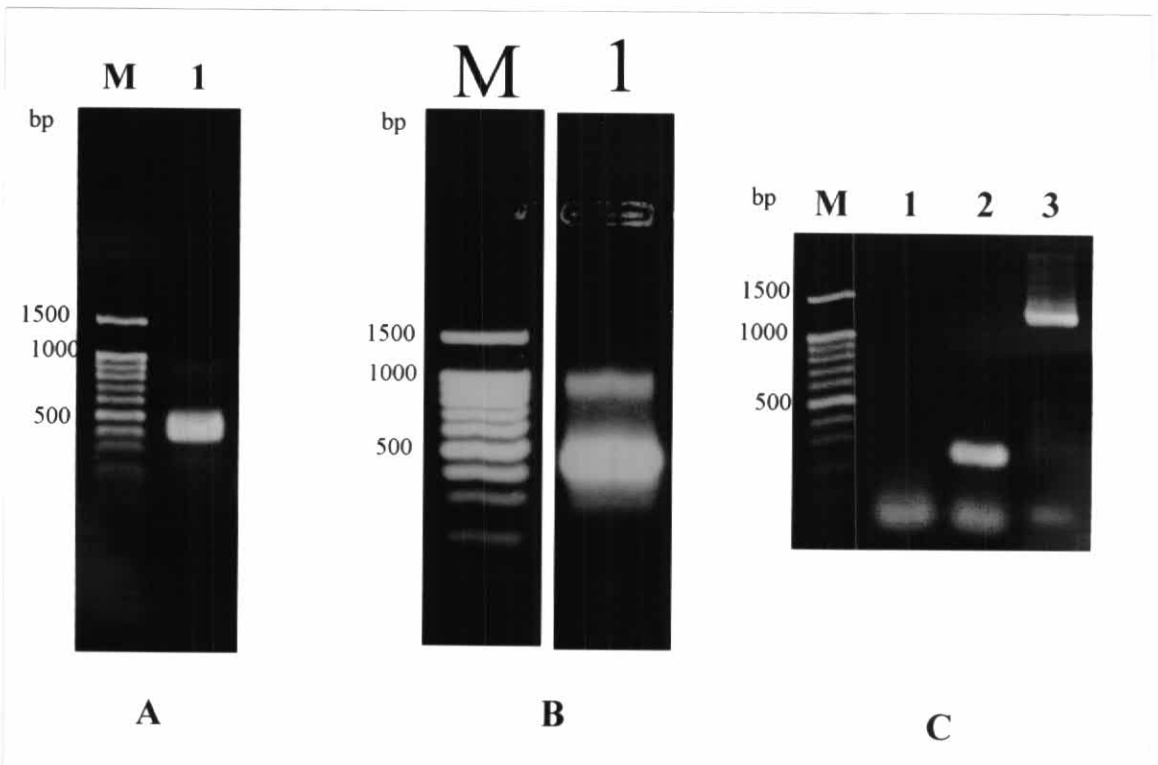


Figure 4.28 RACE PCR products of *chg*-L. 5' and 3' RACE PCR products of *chg*-L are shown in (A) and (B), full length (ORF) PCR product of *chg*-L is shown in (C). Lane M indicates 100 bp DNA markers, lane 1 of (A) and (B) is the products, lane 1, 2, and 3 of (C) is negative control, positive control and the products, respectively.

```
5'L500/12       ACGCGGGGACAGCACCTTGAGAATCTCTCAGATCGCTTGTCACTGTGGAGCCATGGTGAT
5'L500/18       ACGCGGGGACAGCACCTTGAGAATCTCTCAGATCGCTTGTCACTGTGGAGCCATGGTGAT
                ************************************************************

5'L500/12       GAAGTGGACTGCTGCTTGCCTTGTGGCACTGGCTCTATTTGCCAGCGTCTGTGATGCTCA
5'L500/18       GAAGTGGACTGCTGCTTGCCTTGTGGCACTGGCTCTATTTGCCAGCGTCTGTGATGCTCA
                ************************************************************

5'L500/12       GTGGGGAGAGTACACGCCTTCAAAATATCAGAAACCTGCACCTCCTGTGAAGCAAGAGCC
5'L500/18       GTGGGGAGAGTACACGCCTTCAAAATATCAGAAACCTGCACCTCCTGTGAAGCAAGAGCC
                ************************************************************

5'L500/12       CAAACAAGTGCCTCAAGACACTCAACAGCATAAGCAGACATTTGAAACACCACTTCAATG
5'L500/18       CAAACAAGTGCCTCAAGACACTCAACAGCATAAGCAGACATTTGAAACACCACTTCAATG
                ************************************************************

5'L500/12       GACATACCCTGAACCTCCCCCGCCTGAACCTGCGCCTGAAATACCATTTGAACCGCTACG
5'L500/18       GACATACCCTGAACCTCCCCCGCCTGAACCTGCGCCTGAAATACCATTTGAACCGCTACG
                ************************************************************

5'L500/12       TCCTCAACCTGTTGCATCTGTTGCTGTTGAGTGCAGAGAGAATGATGCTCATGTGGAAGT
5'L500/18       TCCTCAACCTGTTGCATCTGTTGCTGTTGAGTGCAGAGAGAATGATGCTCATGTGGAAGT
                ************************************************************

5'L500/12       CAGGAGGGATATGTTTGGG-------------------------------------------
5'L500/18       CAGGAGGGATATGTTTGGGACTGGCCAGTTGGTCAATCCGAATGACCTCACCCTGGGGAA
                *******************

5'L500/12       ----
5'L500/18       CTGT
```

Figure 4.29 Alignments of 5'L500/12 and 18 sequences. Yellow label is start codon, blue label is *chg*-L/D12 R2 primer, and green label is *chg*-L/D12 R1 primer.

From result in Fig. 4.29 found 5'L500/12 and 18 is same sequence but amplified from different primer and shown length of 5'UTR of *chg*-L was 52 bp. Figure 4.30 shown in 100 % identical in coding sequences and some different in 3'UTR of 2 variant of 3'cDNA end cDNA sequences of *chg*-L, variant 19 have longer 3'UTR and polyA than variant 37, variant 19 contained polyadenylation signal but not found in variant 37.

```
3'L500-19       GGCTGGGGAGGTGGAGCTCATCCTAATACTCAAGAAAGGAAGATCCGTGATGTGTCTAAA
3'L500-37       GGCTGGGGAGGTGGAGCTCATCCTAATACTCAAGAAAGGAAGATCCGTGATGTGTCTAAA
                ************************************************************

3'L500-19       TCTCAAACTTTGGAATGGGAAGGTCAAATCACCGTGGGTCCCATTGACATCAAAGAGAGG
3'L500-37       TCTCAAACTTTGGAATGGGAAGGTCAAATCACCGTGGGTCCCATTGACATCAAAGAGAGG
                ************************************************************

3'L500-19       GTACTCGCTTAACACGGTTGTCCATGACCATCACAGCAAACATTAGCCTTCATTCATTAC
3'L500-37       GTACTCGCTTAACACGGTTGTCCATGACCATCACAGCAAACATTAGCCTTCATTCATTAC
                ************************************************************

3'L500-19       TGTCTGTATGTTATTCTCTTGCTACTTGTTGGCAGTTTGATCTACACAAGGAAGGAAAAA
3'L500-37       TGTCTGTATGTTATTCTCTTGCTACTTGTTGGCAGTTTGATCTACACAAGGAAGGAAAAA
                ************************************************************

3'L500-19       GAAAACATTTACTCCATTAAACACAATTGAAAAA-CTGGAAAAGGATGTCATAGCTATCA
3'L500-37       GAAAACATTTACTCCATTAAACACAATTGAAAAAACTGAAAAAAAAAAAAAAAAAAAAAA
                ********************************** *** **** *  * * *  * *

3'L500-19       CCTTCCCCTCGGTGTAATAATAAATGTTACATGTGTACATGTAAAAAAAAAAAAAAAAAA
3'L500-37       AAAAAA------------------------------------------------------

3'L500-19       AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
3'L500-37       ------------------------------------------------------------

3'L500-19       AAAAAAAAAAAA
3'L500-37       ------------
```

Figure 4.30 Alignments of 3'L500/19 and 37 sequences. Yellow label is *chg*-L F-RACE primer, blue label is stop codon, blue alphabet is polyadenylation signal and green label is polyA tail.

Full length (ORF) sequence of *chg*-L were amplified by PCR using primers designed from start and stop codon at 5' and 3' regions of *chg*-L. The product was obtained at the size of 1.3 kb (Fig. 4.27 C). Sequencing (see in appendix B) and BLAST analysis confirmed that this PCR product was *chg*-L which was the closest homologous with chorion protein of gilthead seabream (*Sparus aurata*) with E-value as 7e-145 (see detail in appendix C). Chg-L deduced sequence encoded from ORF was shown in Fig. 4.31. *L. subviridis* Chg-L which contained 419 amino acid residues with molecular weight at 46.09 kDa contained signal peptide at position 1-22, proline-rich region at position 35-86, zona pellucida (ZP) domain (ZP_2) at position 93-353 (score = 35.841) which containing 261 amino acid residues, 8 conserved cysteines residues, and conserved N-glycosylation site at position 193-196 as shown in Fig. 4.31. Consensus pattern domain and motifs show in table 4.3.

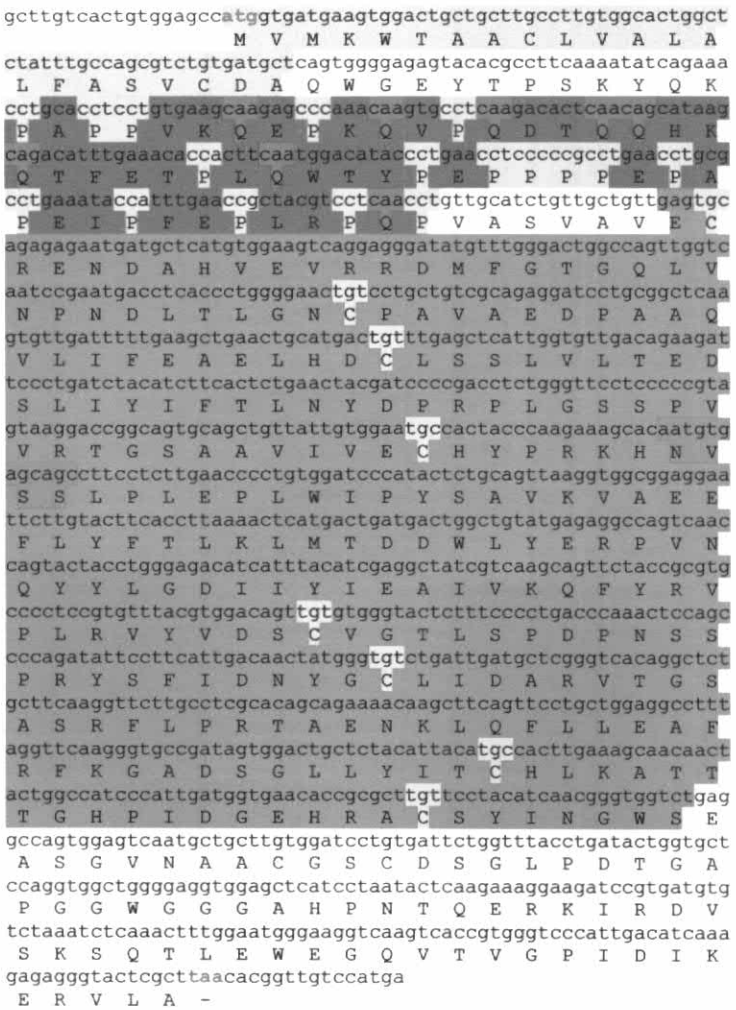
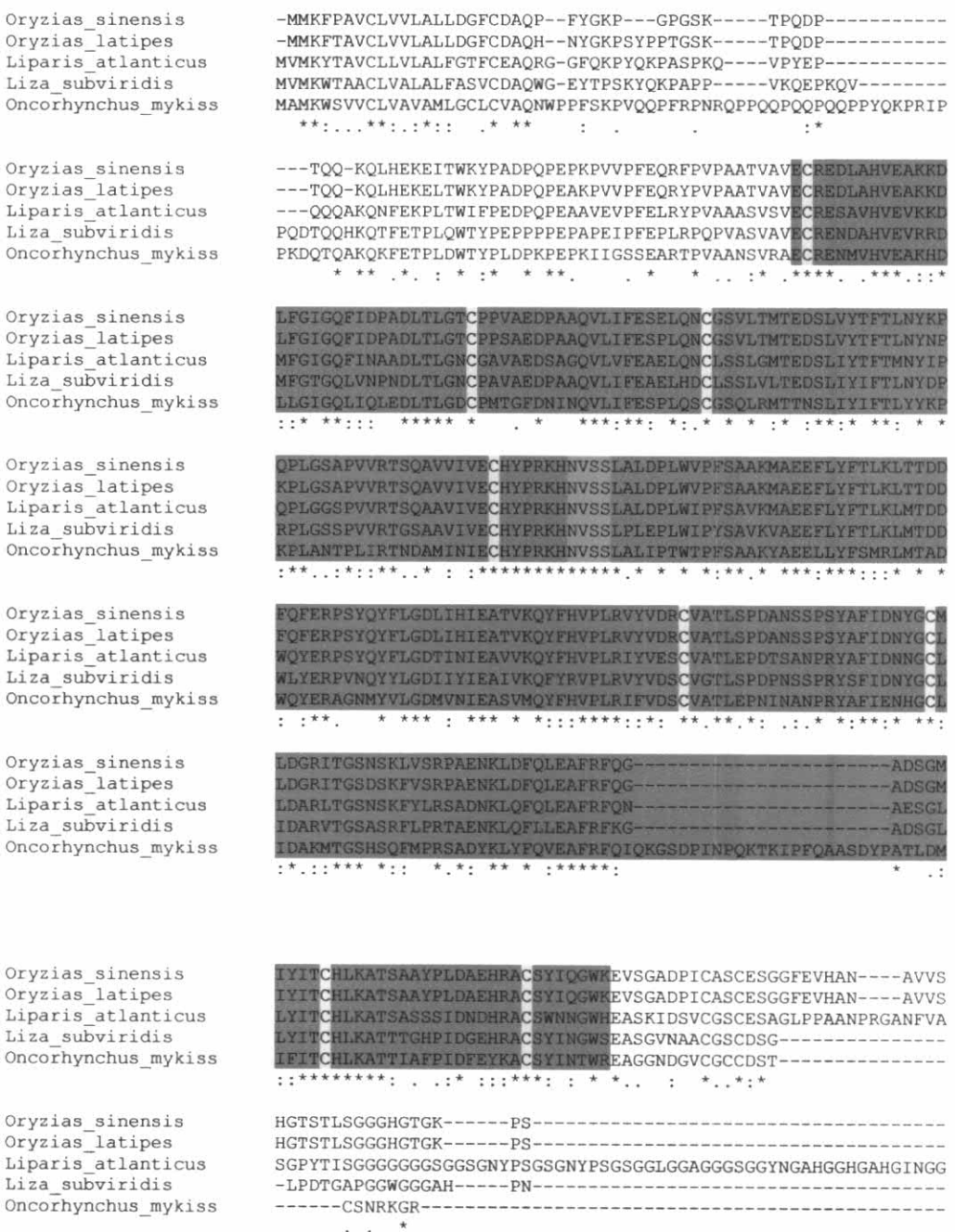


Figure 4.31 Chg-L deduced sequence encoded from chg-L full length sequence. Red and blue alphabet are start and stop codon, grey highlight indicates signal peptide, blue highlight indicates zona pellucida (ZP) domain (ZP_2), green highlight indicates Proline-rich region, and pink highlight indicates conserved N-glycosylation site.

Table 4.3 Domains, motifs, and consensus patterns of Chg-L

Domain	Site	Consensus pattern
zona pellucida (ZP) domain (ZP_2)	93-353	[LIVMFYW] - x(7) - [STAPDNLR] - x(3) - [LIVMFYW] - x - [LIVMFYW] - x - [LIVMFYW] - x(2) - C - [LIVMFYW] - x - [STA] - [PSLT] - x(2,4) - [DENSG] - x - [STADNQLFM] - x(6) - [LIVM](2) - x(3,4) - C The 2 C's may be involved in disulfide bonds

L. subviridis Chg-L was aligned with Chg-L of other fish species as shown in Fig. 4.32, Zona pellucida (ZP) domain of *L. subviridis* Chg-L was highly conserved with this domain of other fish species. ZP domain of Chg-L contained strictly conserved 8 cysteines residues which involved in disulfide bond formation of this protein and one N-glycosylation site. N and C terminus of Chg-L (non label) do not conserved between species of fish.



```
Oryzias_sinensis          ------------------------DPSRKTREAAKTEVLEWEGDVTLGPIPIEERRV-
Oryzias_latipes           ------------------------DPSRKTREAAKTEVLEWEGDVTLGPIPIEERRV-
Liparis_atlanticus        GGGGGGGGGGGGGGYNGNSGGGIYNPSRKIRDVS--EVFEWEGDVTLGPIPIADKMVA
Liza_subviridis           ------------------------TQERKIRDVSKSQTLEWEGQVTVGPIDIKERVLA
Oncorhynchus_mykiss       ------------------------DTTKHQKLVN-----IWEGDVQLGPIFISEKVAQ
                                                     :: : .       ***:* :*** * ::
```

Figure 4.32 Alignments of Chg-L of *L. subviridis* and other species of fish. Zona pellucida (ZP) domain is labeled with green which contained 8 strictly conserved cysteines residues (label with yellow) and 1 conserved N-glycosylation site (label with pink).

An attempt to amplify 5'cDNA end sequence of *chg*-H was not successful. Non-specific bands and smear were obtained from 5'RACE PCR amplification. The result was shown in Fig. 4.34. After sequencing (see in appendix B) and BLAST analysis, this sequence was homologous with chorion protein (*Liparis atlanticus*) with E-value as 5e-62.

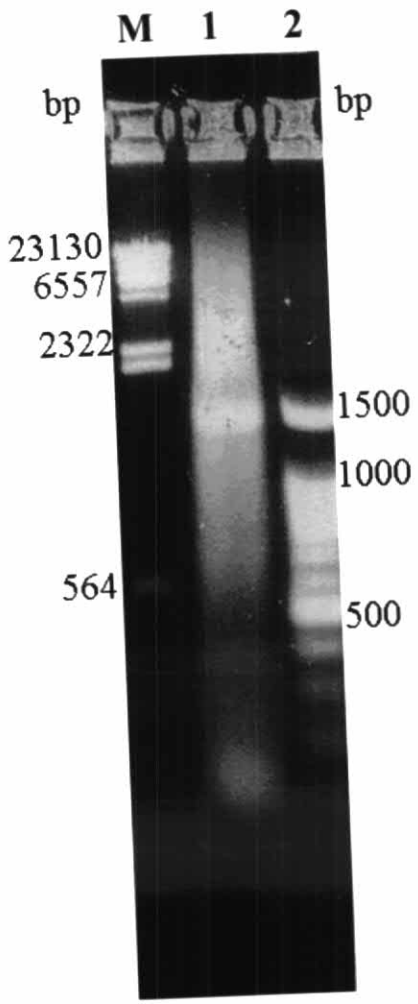


Figure 4.33 5'RACE PCR product of *chg*-H separated in 1.2 % agarose. Lane M1, M2 and 1 is λ*Hin*dIII, 100 bp DNA marker and 5'RACE PCR product of *chg*-H, respectively.

This sequence aligned with complete 5'cDNA end sequence of *chg*-L as shown in Fig. 4.34.

```
5'L500-12            ACGCGGGGACAGCACCTTGAGAATCTCTCAGATCGCTTGTCACTGTGGAGCCATGGTGAT
5'H1.5-10_chg-L_     ------------------------------------------------------------

5'L500-12            GAAGTGGACTGCTGCTTGCCTTGTGGCACTGGCTCTATTTGCCAGCGTCTGTGATGCTCA
5'H1.5-10_chg-L_     ------------------------------------------------------------

5'L500-12            GTGGGGAGAGTACACG-CCTTCAAAATATCAGAAACCTGCACCTCCTGTGAAGCAAGAGC
5'H1.5-10_chg-L_     -----------ACACGACCTTCAAAATATCCTAAACCTGCACCTCCTGTGAAGCAAGAGC
                           ***** *************  ****************************

5'L500-12            CCAAACAAGTGCCTCAAGACACTCAACAGCATAAGCAGACATTTGAAACACCACTTCAAT
5'H1.5-10_chg-L_     CCAAACAAGTGCCTCAAGACACTCAACAGCATAAGCAGACATTTGAAACACCACTTCAAT
                     ************************************************************

5'L500-12            GGACATACCCTGAACCTCCCCCGCCTGAACCTGCGCCTGAAATACCATTTGAACCGCTAC
5'H1.5-10_chg-L_     GGACATACCCTGAACCTCCCCCGCCTGAACCTGCGCCTGAAATACCATTTGAACCGCTAC
                     ************************************************************

5'L500-12            GTCCTCAACCTGTTGCATCTGTTGCTGTTGAGTGCAGAGAGAATGATGCTCATGTGGAAG
5'H1.5-10_chg-L_     GTCCTCAACCTGTTGCATCTGTTGCTGTTGAGCGCAAAGAGAATGATGCTCATGTGGAAG
                     ********************************* *** ***********************

5'L500-12            TCAGGAGGGATATGTTTGGG----------------------------------------
5'H1.5-10_chg-L_     TCAGGAGGGATATGTTTGGGACTGGCCAGTTGGTCAATCCGAATGACCTCACCCTGGGGA
                     ********************

5'L500-12            ------------------------------------------------------------
5'H1.5-10_chg-L_     ACTGTCCTGCTGTCGCAAAGGATCCTGCAGCTCAAGTGTTGATTTTTGAAGCTGAACTGC

5'L500-12            ------------------------------------------------------------
5'H1.5-10_chg-L_     ATGACTGTTTGAGCTCATTGGTGTTGACAGAAGATTCCCTGATCTGCATCTTCACTCTGA

5'L500-12            ------------------------------------------------------------
5'H1.5-10_chg-L_     ACTACGATCCCCGACCTCTGGGTTCCTCCCCCGTAATAAGGACCGGCAGTGCAGCTGTTA

5'L500-12            ------------------------------------------------------------
5'H1.5-10_chg-L_     TTGTGGAATGCCACTACCCAAGAAAGCACAATGTGAGCAGCCTTCCTCTTGAACCCCTGT

5'L500-12            ------------------------------------------------------------
5'H1.5-10_chg-L_     GGATCCCATACTCTGCAGTTAAAGTGGCGGAGGAATTCTTGTACTTCACCTTAAAACTCA

5'L500-12            ------------------------------------------------------------
5'H1.5-10_chg-L_     TGACTGATGACTGGCTGTATGATAGGCCAGTCATACCCGCACTACCTGGGAGACATCATT

5'L500-12            ------------------------------------------------------------
5'H1.5-10_chg-L_     TACATCAGGGCATTCGTCCAAGCAATTTTTCCACCTGCTCCCTCTGTGGTTACTCGTAAA

5'L500-12            --------------
5'H1.5-10_chg-L_     GATGGGTGAGTAAT
```

Figure 4.34 alignments of 5'L 500/12 and 5'H 1.5/10 (*chg*-L). Blue is start codon.

These two sequences were high similarity but 5'H 1.5/10 (*chg*-L) sequence lack overall 5'UTR and some portion of 5'end coding sequence as shown in Fig. 4.34.

```
5'L500-12             MVMKWTAACLVALALFASVCDAQWGEYTPSKYQKPAPPVKQEPKQVPQDTQQHKQTFETP
5'H1.5-10_chg-L_      -------------------------TRPSKYPKPAPPVKQEPKQVPQDTQQHKQTFETP
                                               **** **************************

5'L500-12             LQWTYPEPPPPEPAPEIPFEPLRPQPVASVAVECRENDAHVEVRRDMFG-----------
5'H1.5-10_chg-L_      LQWTYPEPPPPEPAPEIPFEPLRPQPVASVAVERKENDAHVEVRRDMFGTGQLVNPNDLT
                      ********************************* :***************

5'L500-12             ------------------------------------------------------------
5'H1.5-10_chg-L_      LGNCPAVAKDPAAQVLIFEAELHDCLSSLVLTEDSLICIFTLNYDPRPLGSSPVIRTGSA

5'L500-12             ------------------------------------------------------------
5'H1.5-10_chg-L_      AVIVECHYPRKHNVSSLPLEPLWIPYSAVKVAEEFLYFTLKLMTDDWLYDRPVIPALPGR

5'L500-12             ---------------------------
5'H1.5-10_chg-L_      HHLHQGIRPSNFSTCSLCGYS-RWVSN
```

Figure 4.35 Alignments of 5'L 500/12 and 5'H 1.5/10 deduced amino acid sequence.

Deduced amino acid sequence of 5'L 500/12 and 5'H 1.5/10 were high similarity as shown in Fig. 4.35.

Structure of full length of *chg*-L gene which shown position of nucleotide sequence of *chg*-L gene obtained from this study shown in Fig. 4.36.

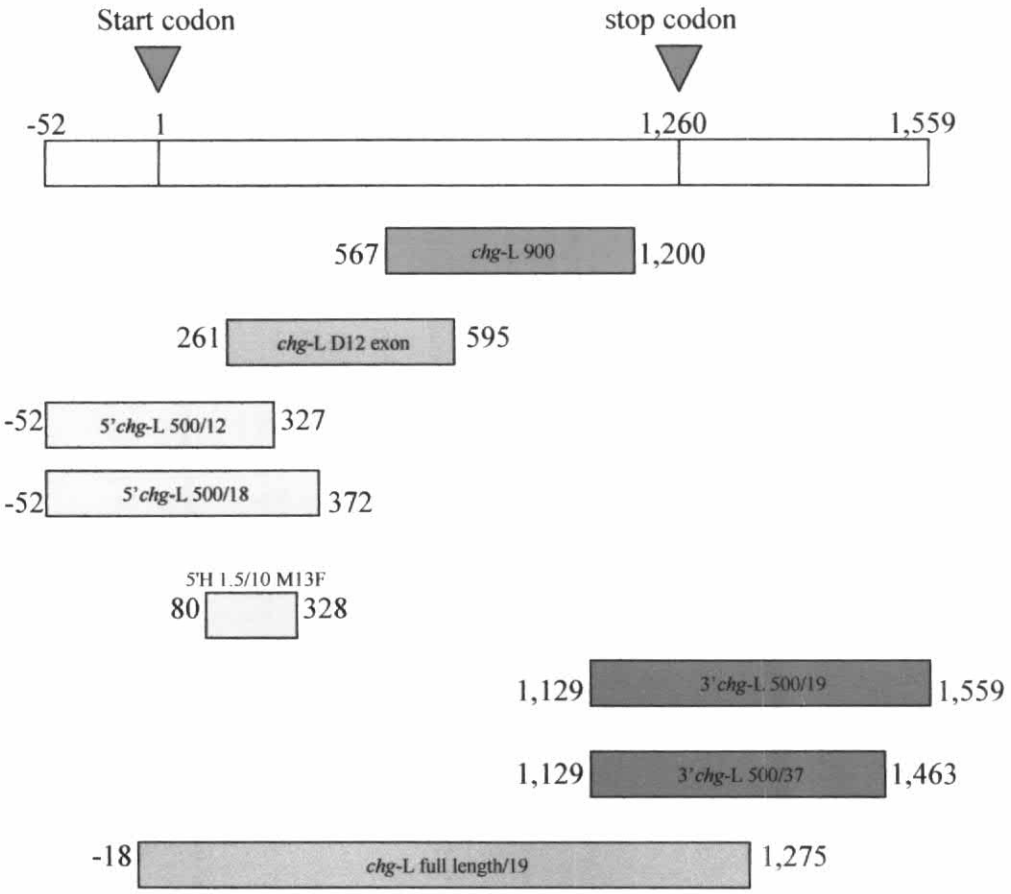


Figure 4.36 Schematic representation of the full length of *chg*-L gene. Blue box is partial coding sequence of *chg*-L, orange box is exon obtained by GenomeWalk PCR from *Dra*I GenomeWalk library, yellow box is 5'end cDNA sequence of *chg*-L, green box is 3'end cDNA sequence of *chg*-L, and purple box is *chg*-L full length (ORF).

4.4.2 choriogenin H (*chg*-H)

chg-H PCR product at the size of 750 bp was amplified from 1st strand cDNA templates of livers as shown in Fig. 4.37 A. After sequencing (see in appendix B) and BLAST analysis, this PCR product was homologous with zona pellucida protein Bb of gilthead seabream (*Sparus aurata*) with E-value as 4e-104. 5'RACE PCR product of *chg*-H at the size of 900 bp and 3'RACE PCR products of *chg*-H at the sizes of 700 bp and 1 kb were amplified from livers (Fig. 4.37 B and C, respectively). 3'RACE PCR product of *chg*-H at size 700 bp and 1 kb were sequenced and BLAST analysis found it homologous with zona pellucida protein Bb of gilthead seabream (*Sparus aurata*) with E-value as 7e-62 and 2e-122, respectively (see detail in appendix B and C).

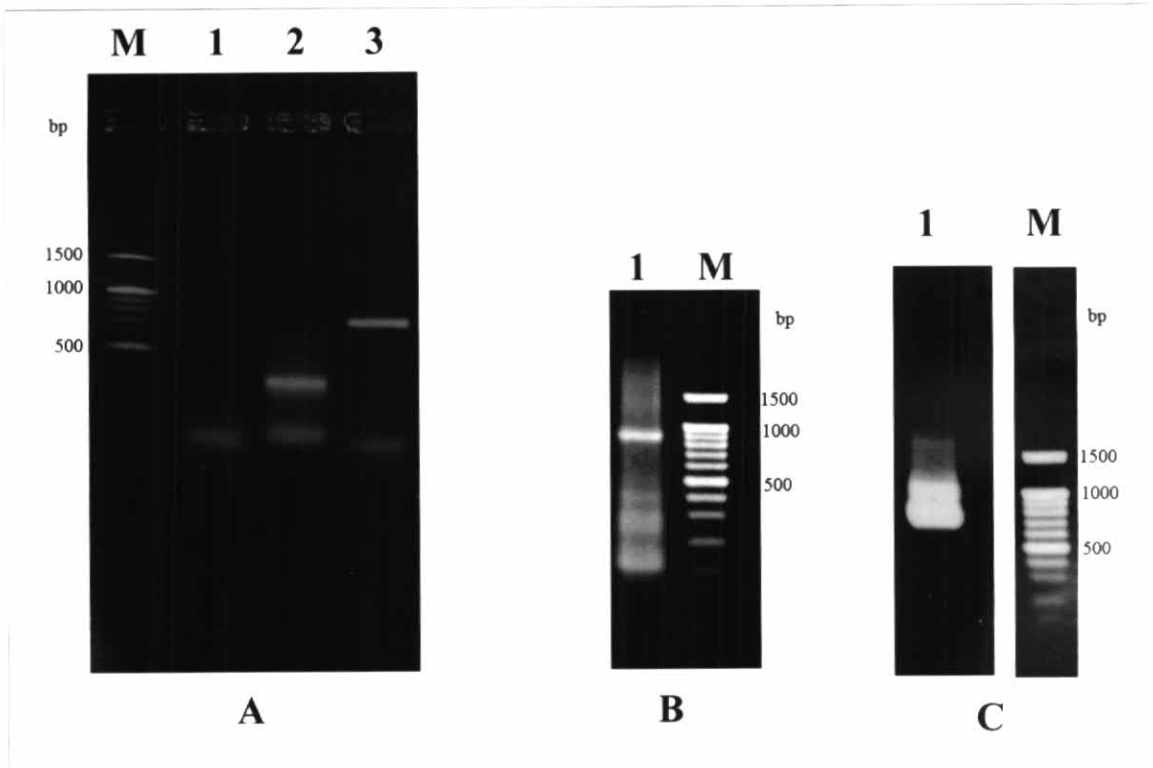


Figure 4.37 *chg*-H PCR products separated in 1.2 % agarose gel. (A) *chg*-H PCR product separated in 1.2 % agarose gel Lane M, 1, 2, and 3 is 100 bp DNA marker, negative control, positive control and chg-H PCR product, respectively, (B) 5' RACE PCR product of *chg*-H separated in 1.2 % agarose gel (lane M and 1 is 100 bp DNA marker and 5' RACE PCR product of *chg*-H, respectively and (C) 3' RACE PCR product of *chg*-H separated in 1.2 % agarose gel (lane M and 1 is 100 bp DNA marker and 3'RACE PCR product of *chg*-H, respectively.

Two variant of 3'end cDNA sequence of chg-H were aligned as shown in Fig. 4.38.

```
3'chg-H800-51     CTACCCTGTGGCCAAAGTATTGAGGGATCCTGTGTATGTGGAGGTTGAGCTCCTTGAAAT
3'chg-H1.1-13     CTACCCTGTGGCCAAAGTATTGAGGGATCCTGTGTATGTGGAGGTTGAGCTCCTTGAAAT
                  ************************************************************

3'chg-H800-51     GACCGATCCAGCACTTGTCCTGACTCTCGGCCGCTGTTGGACAACAACAACCCCCAATCC
3'chg-H1.1-13     GACAGATCCAGCACTTGTCCTGACTCTCGGCCGCTGTTGGACAACAACAACCCCCAATCC
                  *** ********************************************************

3'chg-H800-51     TCACAGCCTGCCCCAGTGGGACATACTGGTGGACGGATGTCCATACAGGGATGATCGTTA
3'chg-H1.1-13     TCACAGCCTGCCCCAGTGGGACATACTGGTGGACGGATGTCCATACAGGGATGATCGTTA
                  ************************************************************

3'chg-H800-51     CTTGTCTGCATTGGTTCCAGTCACTGGTGTCGACCTCCCAGGTCGCTACAGACGTTTCCT
3'chg-H1.1-13     CTTGTCTGCACTGGTTCCAGTCACTGGTGTCGACCTCCCAGGTCGCTACAGACGTTTCCT
                  ********** ***************************************************

3'chg-H800-51     TTTCAAAATGTTCACCTTTGTGGATCCTGCTTCATTGGAGCCCCTGAGAGAATACGTGTA
3'chg-H1.1-13     TTTCAAAATGTTCACCTTTGTGGATCCTGCTTCATTGGAGCCCCTGAGAGAATACGTGTA
                  ************************************************************

3'chg-H800-51     CATTCACTGTAGTACAGCTGTGTGTGCTGCTGCACCAGGCCGTAACTGTGATCCATCATG
3'chg-H1.1-13     CATTCACTGTAGTACAGCTGTGTGTGCTGCTGCACCAGGCCGTAACTGTGATCCATCATG
                  ************************************************************


3'chg-H800-51     CTACAGGAAAAAGAGATCTGTTGATGCCGTGGTCCAGAGAAAGGATGAACCCAAGGTTGT
3'chg-H1.1-13     CTACAGGAAAAAGAGATCTGTTGATGCCGTGGTCCAGAGAAAGGATGAACCCAAGGTTGT
                  ************************************************************

3'chg-H800-51     TGTTTCTTTTGGACCAGTGATCATGGCCGCCCCTGAGGAGCAACAGGCTGAAGAATAGAG
3'chg-H1.1-13     TGTTTCTTTTGGACCAGTGATCATGGCCGCCCCTGAGGAGCAACAGGCTGAAGAATAGAG
                  ************************************************************

3'chg-H800-51     CTGAGAACATTTTGAAAAAATATTTGGAACATAGTACAGTTGTTAGTTGTATGAAATTAA
3'chg-H1.1-13     CTGAGAACATTTTGAAGAAATATTTGGAACATAGTACAGTTGTTAGTTGTATGAAATTAA
                  *************** ********************************************

3'chg-H800-51     AATGTCAATGTACGTATTTGTACATTTTGCAGAAAGCCCCGAGAGAACATGAAAACTGCA
3'chg-H1.1-13     AATGTCAATGTACGTATTTGTACATTTTGCAGAAAGCCCCGAGAGAACATGAAAACTGCA
                  ************************************************************

3'chg-H800-51     TGTGTTGTAAAAACCCAACTCGCGAAACGCTCATATTATAAAATTTAAATGAATAAATAC
3'chg-H1.1-13     TGTGTTGTAAAAACCCAACTCGCGAAACGCTCATATTATAAAATTTAAATGAATAAATCC
                  ************************************************************ *

3'chg-H800-51     ATTGAAAGCTCAAAAAAAAAAAAAAAAAAAAAAAAAA----------------------
3'chg-H1.1-13     ATTGAAAGCTTACAAGGATTTGCTTCATGATATTATTCATAAGTCAGCAAGCTGCCATTT
                  ********** * **  *        *  * *  *   *

3'chg-H800-51     ------------------------------------------------------------
3'chg-H1.1-13     TAACTAGTAATGTTAGCCGGTGATATGGGAGTGGATTGCCTTGTAATTGAGGGGTTTCTG


3'chg-H800-51     ------------------------------------------------------------
3'chg-H1.1-13     TTGTCAGTATTTGTTTATTCAAACACACACACACATTTTTCAGTATAACTAAATTTAGCA


3'chg-H800-51     ------------------------------------------------------------
3'chg-H1.1-13     TAATACTACAGCAAGTTACAGCACTGTATATGCCCCATGGATTCACGTGAACTATTGCAA
```

```
3'chg-H800-51      ------------------------------------------------------------
3'chg-H1.1-13      CAAAACGGGTCACTTAAAATGTTCATGCTTCTCAAATAAACTATATACGGGCAAAAAAAA

3'chg-H800-51      ----------
3'chg-H1.1-13      AAAAAAAAA
```

Figure 4.38 Alignment of 3' end cDNA sequence of *chg*-H. Green label is stop codon, yellow label is polyadenylation signal and blue label is polyA tail.

Fig. 4.38 has shown different in 2 nucleotides position of coding sequence between 2 variant and high similarity in 3' UTR of 2 variant but 3'*chg*-H 1.1-13 have 3' UTR longer than 3'*chg*-H 800-51. Coding sequence of 2 variant was translated to amino acid sequence, 2 variant were 100 % identical in amino acid sequence as shown in Fig. 4.39.

```
3'Chg-H800-51      YPVAKVLRDPVYVEVELLEMTDPALVLTLGRCWTTTTPNPHSLPQWDILVDGCPYRDDRY
3'Chg-H1.1-13      YPVAKVLRDPVYVEVELLEMTDPALVLTLGRCWTTTTPNPHSLPQWDILVDGCPYRDDRY
                   ************************************************************

3'Chg-H800-51      LSALVPVTGVDLPGRYRRFLFKMFTFVDPASLEPLREYVYIHCSTAVCAAAPGRNCDPSC
3'Chg-H1.1-13      LSALVPVTGVDLPGRYRRFLFKMFTFVDPASLEPLREYVYIHCSTAVCAAAPGRNCDPSC
                   ************************************************************

3'Chg-H800-51      YRKKRSVDAVVQRKDEPKVVVSFGPVIMAAPEEQQAEE
3'Chg-H1.1-13      YRKKRSVDAVVQRKDEPKVVVSFGPVIMAAPEEQQAEE
                   ***************************************
```

Figure 4.39 Alignment of 3'Chg-H 800/51 and 1.1/13 deduced sequence.

Partial and 3' end cDNA sequence were ligated *in silico* and translated as shown in Fig. 4.40. It contained 2 polyadenylation signals.

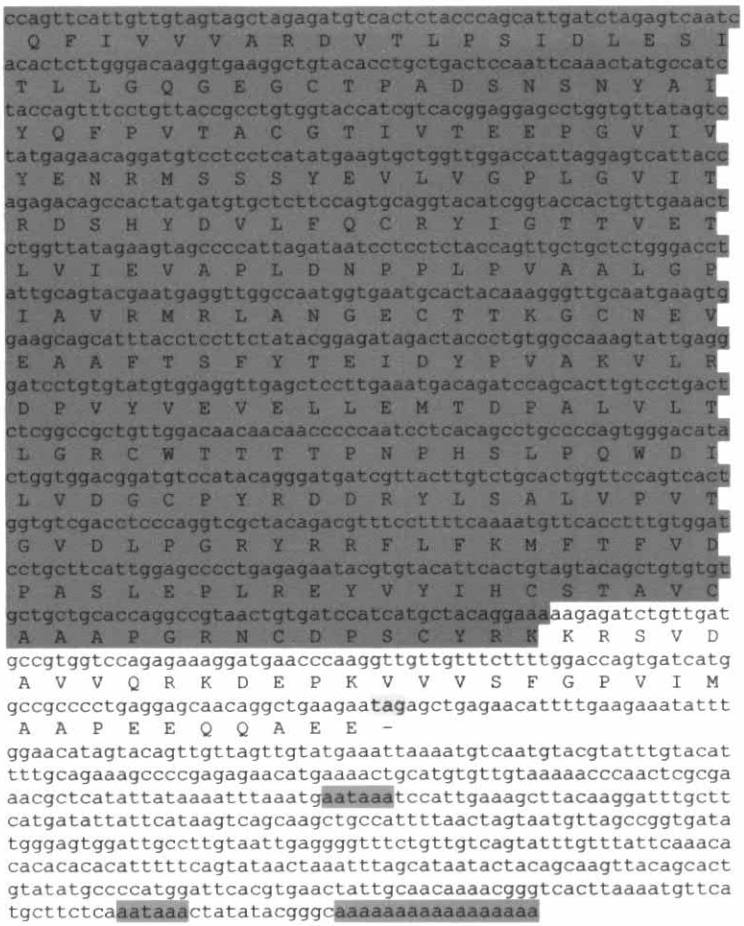


Figure 4.40 Deduce sequence of Chg-H. Green label is ZP domain, yellow label is stop codon, blue label is polyadenylation signal, and pink label is polyA tail.

Partial Chg-H contained ZP domain at position 1-275 was aligned with other fish Chg-H as shown in Fig. 4.41.

```
Liza_subviridis         ------------------------------------------------------------
Oryzias_latipes         MARHWSITVFSALALLCSFLGTEVDAQKGNPQDPKVPYPPYYPQPKPQDPQHVSPPYYPG
Liparis_atlanticus      ------------------------------------------------------------
Oncorhynchus_mykiss     --MKWSAVCLVAVATLGWLCDAQIYLEK--PGWPPIQTPPSWPAQPPQ------------

Liza_subviridis         ------------------------------------------------------------
Oryzias_latipes         KPQNPPQKPSNPQYPSYPQTPQNPQVPQNPQVPQNPQYPSYPQNPSYPQNPSYPQYPSNP
Liparis_atlanticus      ------------------------------------------------------------
Oncorhynchus_mykiss     KPIQPPQRPAQP--------------------------------------------PQWPVQP


Liza_subviridis         ------------------------------------------------------------
Oryzias_latipes         PTSQNPSYPQNPKLFQDGKPSNPQQPQVPQYPSKPQPPQNPQVPQYPSKPQPPQNPQVPQ
Liparis_atlanticus      ------------------------------------------------------------
Oncorhynchus_mykiss     P----------------------------QRPAQPPQRP------------------
```

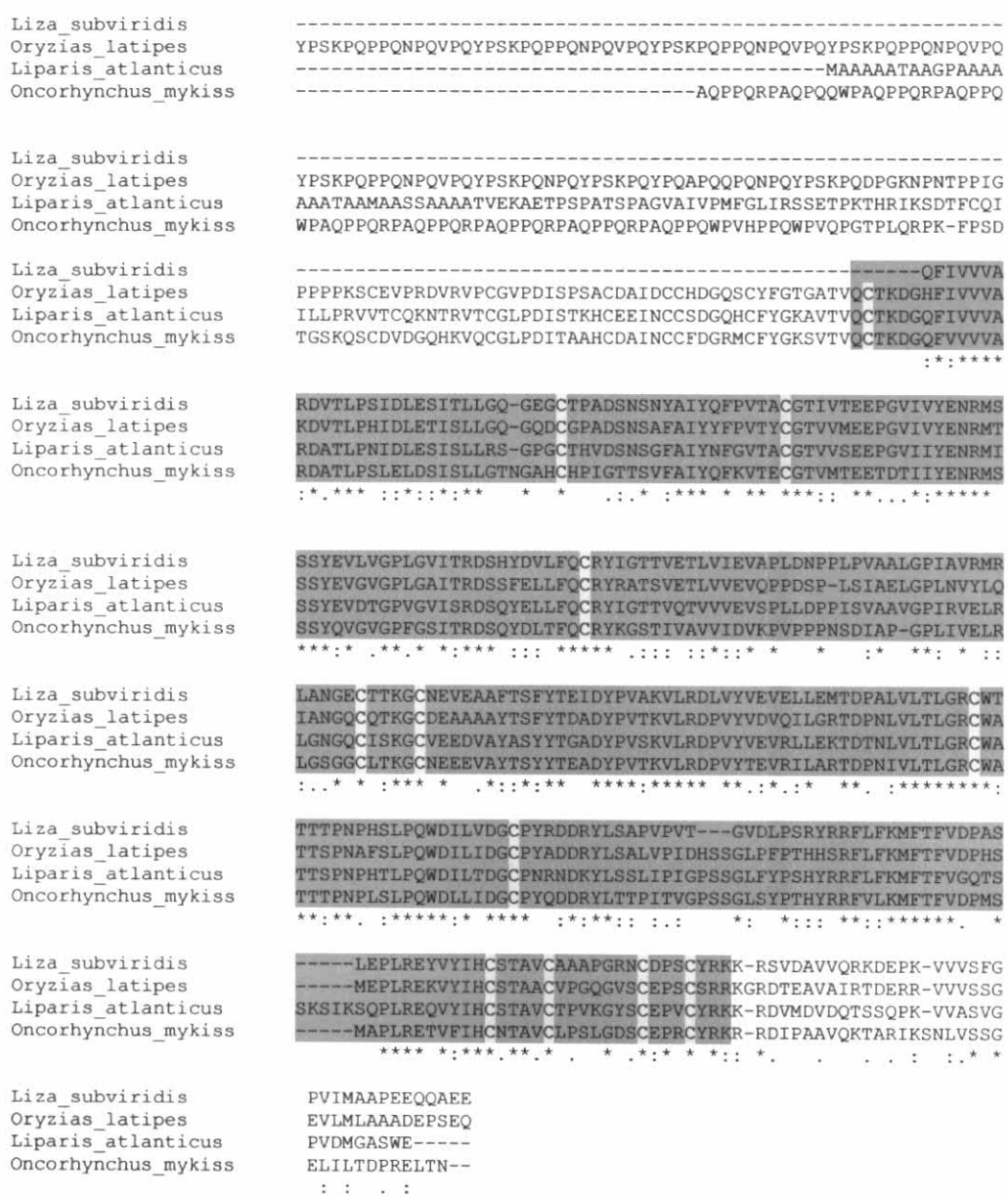


Figure 4.41 *L. subviridis* Chg-H aligned with other fish Chg-H. Blue label is zona pellucida (ZP) domain, conserved cysteines residues label with yellow.

ZP domain of Chg-H was conserved and contained 12 conserved cysteines residues as shown in Fig. 4.41.

4.5 Cloning and characterization of vitellogenin (*vtg*) genes

4.5.1 *vtg*-1

Partial cDNA sequence of *vtg*-1 at the size of 400 bp was amplified from 1st strand cDNA templates of liver (Fig. 4.42 A) The results of sequencing (see in appendix B) and BLAST analysis indicated its close similarity with *vtg*1 of white mullet (*Mugil curema*) with E-value as 9e-114. 5' RACE PCR product at the size of 250 bp (Fig. 4.42 B) and 3' RACE PCR product at the sizes of 4.3, 1.5, 1.1, 0.6 and 0.5 kb (Fig. 4.42 C) were obtained. Only 250 bp product from 5' RACE PCR and 4.3 kb product from 3' RACE PCR were cloned and sequenced (see detail in appendix B). The result of sequencing and BLAST analysis indicated that both sequences were homologous with *vtg*A of charr (*Salvelinus alpinus alpinus*) and *vtg* of red seabream (*Pagrus major*) with E-value as 1e-04 and 0.0, respectively. Because 3'*vtg*-1 sequence was very long, only 75% of the 4.3/51 fragment was sequenced. Nucleotide sequence of *vtg*-1 (*in silico* ligated 5', 3' end cDNA sequence and partial coding sequence) shown in Fig. 4.43. Because internal sequence at size 950 bp not sequenced so sequence divided to 2 parts and translated to amino acid sequence of N and C-terminus of Vtg-1 as shown in Fig. 4.44 (A) and (B), respectively. *L. subviridis* Vtg-1 which contained 1,550 amino acid residues with molecular weight at 170.5 kDa which contained lipoprotein N-terminal domain (LPD_N) at position 24-589, lipovitellin-phosvitin complex; beta-sheet shell regions at position 616-888 and von Willebrand factor (vWF) type D domain (VWD) at position 1,280-1,435 (Fig. 4.44). LPD_N, beta-sheet shell regions, and VWD of *L. subviridis* Vtg-1 were aligned with these domains of other fish species (Fig. 4.45 (A), (B), and (C), respectively). LPD_N, beta-sheet shell regions, and VWD of *L. subviridis* Vtg-1 were highly conserved with these domains of other fish species.

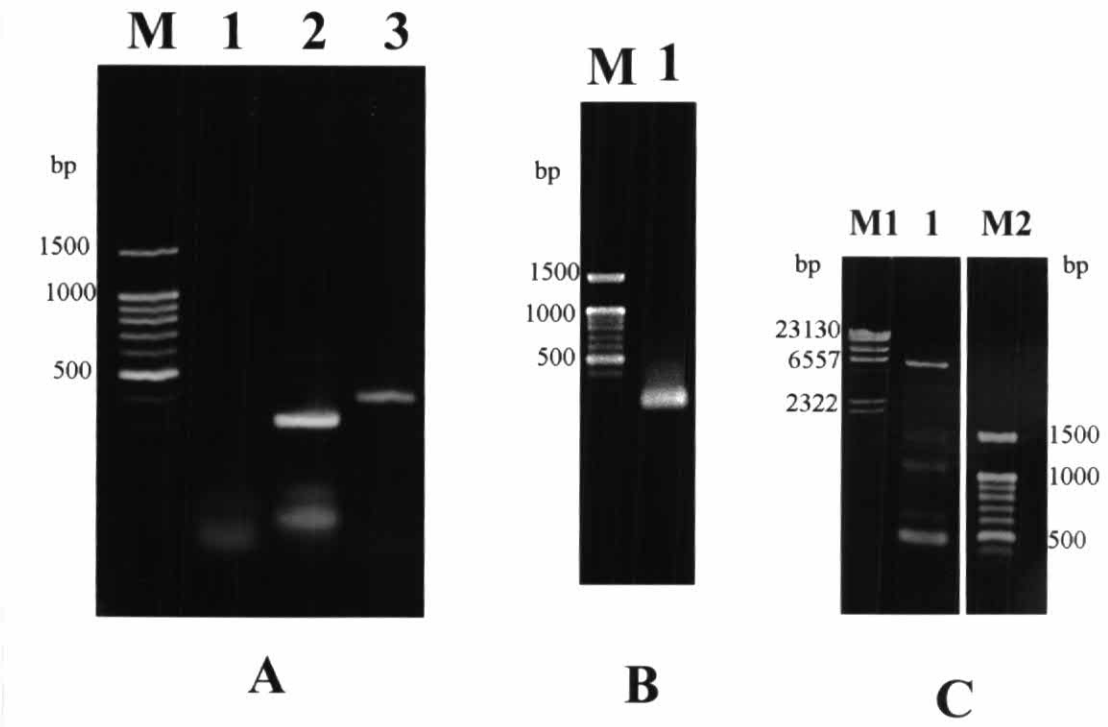


Figure 4.42 *vtg*-1 PCR products separated in 1.2% agarose gel. (A) *vtg*-1 PCR product separated in 1.2% agarose gel, lane M, 1, 2 and 3 is 100 bp DNA marker, negative control, positive control and *vtg*-1 PCR product, respectively, (B) 5' RACE PCR of *vtg*-1 product separated in 1.2% agarose gel. Lane M and 1 is 100 bp DNA marker and 5'RACE PCR of *vtg*-1, respectively and (C) 3' RACE PCR of *vtg*-1 separated in 1.2% agarose gel .Lane M1, M2 and 1 is λ*Hin*dIII, 100 bp DNA marker and 3' RACE PCR of *vtg*-1, respectively.

ACGCGGGAAGCAGTGCTATCAACGCACAGTACGCGGGGACAACACCATCGTCC**ATG**ACAGCACTTGTACTTGCCCGGACACTGG
CCTTCATGGCTGGCCAGACGCAACATTTTGCCCCTGAATTTGCCGCCAGTAAAACATTTGTCTACCACCTAGGAGGCATTGCTC
CCGGGGGGTCTGTCAGAGGAAGGTTTGGCGAGAGCTGGACTCAAAGTCAGCAGCAAAGTTCTCATCAGCGCTGCAGCTGAAAAC
AAATACCTGCTGAAGCTTGTGGAACCTGAGCTCTATGAATACAGTGGTGTTTGGCCCAAGGATCCTTTAATCCCAGCAGCCAAG
CTGACTTCAGCCCTGGCAGCTCAGCTTGTGACTCCCATCAAGTTTGAATATGCTAATGGTGTGGTAGGGAAACTGCTCGTCCCT
GAAGGAGTCTCAACAATGGTGCTGAACATCCACAGAGGCATCCTGAATCTCCTTCAGCTCAACATCAAGAAGACTCAGAATGTA
TACGAGCTGCAGGAGGCTGGAACTCAGGGAGTGTGCAAGACCCTTTATGCCATCAGTGAAGATGAAAGGGCTGAACGTATCCTT
CTGACAAAGACCAGGGACATGAACCAATGTCAGGAAAAGATCATCAAGGATATGGGGTTGGCGTATACAGAGAAATGTGCCAAG
TGTCAGCAGGATTCCAAAAACCTGAGAGGTGCAACATCATTCAATTATATCTTGAAGCCAGTTGCTAGTGGCGTCCTTATCCTG
GAGGTAGCTGTGAATGAGGTGATCCAGTTCTCACCATTTGCTGAGTTGAAAGGAGCTGCTCAGATGGAAACCAAGCAATCATTG
GTCTTCCTTGAGATTCAGAGAGCCCCCATTGCACCCATTGAGGCTCAGTATATTCATCAAGGATCTCTTAAGTACGAGTTCTCC
ACTGAGCTTCTTCAGACACCCATTCAGCTAATAAAGATCAACAATGCACAGGCCCAGATCGTGGATGTCCTGAATCACCTGGTT
ATCCACAATGCGGAGAGAGTCCATGAGGATGCCCCTCTGAAGTTTTTGGAATTAATTCAGCTCCTCCGTGCAGCCCGCTTTGAA
GACCTAGAAATGACCTGGAACATACAAAAGGATCCCTGCCCACAGACAGTGGATCCTGGATGCTGTCCCAGCCATTGGAAGTGC
TGCTGCTCTACGATTCATCATGGAGAAATTCCTGTCAGATGACATAACTGTTGCTGAAGCAGCTCAGGCTTGATTGCATCTGTT
CACATGGTGACAGCAGACACAGAGGCCATTAAGCTGATTAAGTCCCTGGCAGTCAATAGCAAAGTAATCGACAACCCAGTTCTG
CGTGAGATTGTCCTCCTTGGGTATGGCACCATGATTTCCAAACACTGCGTTGAGTTGGCTGTCTGCCCGGCTGAGCTCATAAAG
CCTATCCAGGACCTTCTTGCTGAGGCTGTCACCAAGGACGAGACAGAGGACATCATTCTGCTTCTGAAGGTTATGGGAAATGCT
GGACATCTTGCTAGCCTCAAGGCAATCACAAAGATCCTGCCCATACATGGTACTGCAGCTGCAGTTCTGCCAGTGAGAGTTCAT
GTTCAAGCAATCATGGCCCTGAGGAACATTGCAAAGAAGGAGCCCAGAATGGTCCAGGAACTGGCTCTTCAACTCTATATGGAT
AAGGTTCTTCACCCAGAGCTTCGTATGATCGCATGCATTGTGCTGTTTGAGACAATGCCTCCAATGGGTTTGGTGACAACTGTT
GCCAACATTGTAAAGACTGAAGAGAATCTGCAGGTGGCAAGCTTCACTTACTCTCATATGAAATCCCTCACCAGGAGCAGTGCA
GCTATTCATGCTTCAGTTGCTGCAGCTTGCAACATTGCCGTTAAACTCTTGAGCCCAAGGCTGGACAGACTCAGCTTACGTTTC
AGCAAAGCCATCCATGTGGACGTCTACAACAATCCCTTGATGCTTGGTGCTGCTGCTGGTGCTTTCTACATCAATGATGCTGCC
ACCAACGTTCTGGAGGTTGGAGTGAGAACTGAGGGACTACAGGAGGCTCTTCTGAAGAACCCTGTGCTCATTGACAATGCTGAC
AGGATCACCAAGATGAAGCGTGTCATTAAGGCTCTCTCTCACTGGAGGTCTCTGCCCTCCACCCAACCGCTGGCCTCCGTCTAC
ATCAAGTTCTTTGGACAAGAAATTGCCTTCGCTAACATTGACAAAGCCTTAATTGACCAAGCAATTGCGCTTGCTACTGGACCC
TCTGTTCAGGCATTAGGCAGAAATGCTATCAAGACTCTGCTGTCTGGTGCTTCCTTACACTTTGCTAAGCCTCTGCTCATTACT
GAGATGCGTCGCATTCTGCCAACTGCTGCTGGTCTTCCAATGGAGCTCAGTCTGTACACTGCTGCTGTGGCTGCAGCAGGTGTC
AAAGTCCAAGCAAAAACAGTACCAGCTCTGCCAGAAAACTTCCATTTCGCTCAGCTCCTAAAGACAGACATACAGCTTGAGACC
GAGATCAGACCAAGTGTTGCTATGAACACATTTGCCGTTATGGGTGTAAACACTGCTGTACTGCAAGCTGCTGTGCTATCAAAA
GCTAAACTCAACTCCCTTGTGCCAGCTAAAATTGCTGCAAGACTTGACATCAATGAGGGACACTTTAAGATTGAAGCTCTTCCT
GGAACTGTGCCTGAAAACACTGCAGTTGTCCATGTTGAGACACCTGCTGTTGCAAGA(N)$_{550}$
TTACAGGTTTCAAAAAGAATCACAAGAACATGGCTCGTACTTCTCGGGGTGCCTCTGCAGTGGTTTCCAGAAGCAGGAGCAGTG
CCTCAAGCTTTGAGGCCATCTACAGAAAGAATAAATTCCTAGGGAATGACGCTGCTCCTACCTTTGCCATCATCTTCCGTGCCA
TTAGAGCTGACAACATGATGCAGGGATACCAACTTGCTGCCTATTTGGACAAACCCAGTGCCAGACTCCAGATCATTCTCGCTG
CCTTAGCTGCTGACCTTGTCTTGCAATCAGATTCTTGCACAGGATTGCACAGATGAGATGAAATTCATGGTTCTCCTGAAGAAA
GATCACATCAAGCAGAACCACATCAATGTGAAAATTGCTGACATAGATATTGACCTGTACCCAAAGAACAGTGACGTGGTTGTG
AAGGTTAATGGAATGGAAATGCCCATCAGCAACCTCCCATACCAGCACCCCACAGCCAATATCCAGATCAGATCAAAGGGTGAA
GGCATTGCTATGTACGCATCTAGCCATGGTCTTCATGAAGTCTACTTTGACAAGAACTCATGGAAGATTAAAGTTGTGGACTGG
ATGAAGGGGAAGACATGTGGACTGTGTGGAAAGGCTGATGGGGAGATCAGACAGGAATACCGCACACCCAACGGACGCCTGACT
AAGAATGCAGTCAGCTTTGCTCATTCTTGGATTCTGCCAGCTGAGAGCTGCAGGGACACCACTGGGTGCCGTGTGAAGCTTGAA
TCTGTGCAGCTAGAGAAACAGGTGAACGTGCATGGTCTAGAGTCCAGATGCTACTCTGTTGAGCCTGTCCTTCGCTGTTTGCCT
GGCTGCTTCCCTCTAAAGACCACCAGTGTCACTGTTGGCTTCCACTGTCTACCTGCTGATTCCATCCAGAGTCGCCCTGAGAGT
ATAAGCAACATTTATGAAAACGGCGTGGACCTGAAGGAAACAGCAGAGGCTCACCTTGCCTGCAGCTGCACCGCTCAGTGCGCT
TAATAACATGTCCTGTATGTTATTTTAAATAGACTGCAACTGAAGCTGAAAGTCAAGCAAATGGGTACGAGCCTCCTATTGGAT
GATTACTGCACGGGTGATTATCCTTCAGCTGGCAACAAGTCATTTAACCCCAACTGATGCAATGAGTAGCTTCTCATTTCTTAA
ACATCCATGTTGAAAGACACATCCTGTGTTCATGAAAAAGATGTTAGTTTGAGAAGTCAAATCATTGGCATGCATCAAGCAGAG
AAAACATTTAAGGAGATTGAAGAAACTACCAAAACTGGGTAACGAACTGTCTAACATTATTAAAAACTGGAAGGATAGTGGAGA
ACCATTGTCTTTGAGGAAGAAATGTGGTCAAAAAAAAAAAATCTGATCGTGATAGGCGATCACTTAAATGTTAGGTGAAATCATA
TTGAAGAAAAACAAGTTTTTTATTTTTTTTGGAGGTGACTATTTTTTGGCCAGGCAGTGTATAAAGGACTTTCATATGCATTTG
ATGTAAAGGGAAGCCAGTGTTATGACCATCTTGGATAATAGTATTTGAATGGTCAGATCACAGAAGAAAAACTTGAATTCAGAA
GCACTTGAATTCAGCAATACTACACAGTGTAAATAAGAGTCTAGAGTCTAAAAAAGGTTTTCCTGTCATTTAAATATTTGTGAA
TACTTGATTGTTGACAACAAGAAAATTAAACACCATCTTTGTCG**AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA**

Figure 4.43 Full length mRNA sequence of *vtg*-1. Yellow label is start codon, blue label is non-sequenced sequence in 3'v1 4.3/51 insert, green label is stop codon, and pink label is polyA tail.

```
MTALVLARTLAFMAGQTQHFAPEFAASKTFVYHLEALLPGGLSEEGLARAGLKVSSKVLISAAAENKYLLKLVEPELYEYSGVW
PKDPLIPAAKLTSALAAQLVTPIKFEYANGVVGKLLVPEGVSTMVLNIHRGILNLLQLNIKKTQNVYELQEAGTQGVCKTLYAI
SEDERAERILLTKTRDMNQCQEKIIKDMGLAYTEKCAKCQQDSKNLRGATSFNYILKPVASGVLILEVAVNEVIQFSPFAELKG
AAQMETKQSLVFLEIQRAPIAPIEAQYIHQGSLKYEFSTELLQTPIQLIKINNAQAQIVDVLNHLVIHNAERVHEDAPLKFLEL
IQLLRAARFEDLEMTWNIQKDPCPQTVDPGCCPSHWKCCCSTIHHGEIPVRHNCCSSSGLIASVHMVTADTEAIKLIKSLAVNS
KVIDNPVLREIVLLGYGTMISKHCVELAVCPAELIKPIQDLLAEAVTKDETEDIILLLKVMGNAGHLASLKAITKILPIHGTAA
AVLPVRVHVQAIMALRNIAKKEPRMVQELALQLYMDKVLHPELRMIACIVLFETMPPMGLVTTVANIVKTEENLQVASFTYSHM
KAIHASVAAACNIAVKLLSPRLDRLSLRFSKAIHVDVYNNPLMLGAAAGAFYINDAATNVLEVGVRTEGLQEALLKNPVLIDNA
DRITKMKRVIKALSHWRSLPSTQPLASVYIKFFGQEIAFANIDKALIDQAIALATGPSVQALGRNAIKTLLSGASLHFAKPLLI
TEMRRILPTAAGLPMELSLYTAAVAAAGVKVQAKTVPALPENFHFAQLLKTDIQLETEIRPSVAMNTFAVMGVNTAVLQAAVLS
KAKLNSLVPAKIAARLDINEGHFKIEALPGTVPENTAVVHVETPAVAR
```

A

```
TGFKKNHKNMARTSRGASAVVSRSRSSASSFEAIYRKNKFLGNDAAPTFAIIFRAIRADNMMQGYQLAAYLDKPSARLQIILAA
LALTLSCNQILAQDCTDEMKFMVLLKKDHIKQNHINVKIADIDIDLYPKNSDVVVKVNGMEMPISNLPYQHPTANIQIRSKGEG
IAMYASSHGLHEVYFDKNSWKIKVVDWMKGKTCGLCGKADGEIRQEYRTPNGRLTKNAVSFAHSWILPAESCRDTTGCRVKLES
VQLEKQVNVHGLESRCYSVEPVLRCLPGCFPLKTTSVTVGFHCLPADSIQSRPESISNIYENGVDLKETAEAHLACSCTAQCA
```

B

Figure 4.44 N-terminus amino acid sequences of *L. subviridis* Vtg-1 (A) and C-terminus amino acid sequences of *L. subviridis* Vtg-1 (B). Blue label is lipoprotein N-terminal domain, green label is lipovitellin-phosvitin complex; beta-sheet shell regions and yellow label is von Willebrand factor (vWF) type D domain.

```
Pagrus_major                MRAVVLALTLALVAGQPHNLAPEFAAGRTYVYKYETLLLGGLPEEGLAKA
Sillago_japonica            MRVVALALTLALVAGHPYNLVPEFAAGKTYVYKYEALILGGLPEEGLAKA
Verasper_moseri             MRVVALALTLALVAGHPQNFAPDFAAGKTFLYKYEALLLSGLPEEGLARA
Liza_subviridis             MTALVLARTLAFMAGQTQHFAPEFAASKTFVYHLEALLPGGLSEEGLARA
Melanogrammus_aeglefinus    MRAVVLALALAVVAGQYVNFAPDFASSKTYVYKYEAVLMAGLPEEGLASA
                            * .:.** :**.:**:  ::.*:**:.:*::*: *::: .**.***** *

Pagrus_major                GLKISSKVLISATAENTYMLKLAEPEIFELSGIWPKDPLIPATKLTSALA
Sillago_japonica            GLKVSSKVLISAAAENTYLLKLVEPELFEYSGVWPRDPLVPATKLTSALA
Verasper_moseri             GLKVSSKVLISAAEQNIHMLKLVEPELFEYSGVWPKDPLVPATKLTSALA
Liza_subviridis             GLKVSSKVLISAAAENKYLLKLVEPELYEYSGVWPKDPLIPAAKLTSALA
Melanogrammus_aeglefinus    GLKVLSKVLISNIAENTYLLKLADPELFEYSGVLPKDPFIPANMLKEALA
                            ***: ******   :* ::***.:**::* **: *:**::**  *..***

Pagrus_major                AQLMTPIKFEYTNGVVGKMFAPEGVSTMVLNVYRGILNVLQLNIKKTHNV
Sillago_japonica            AQLMTPIKFEYVNGVVGKMFAPEGISTMVLNVYRGILNVLQLNIKKTQNI
Verasper_moseri             AQLLTPIKFEYINGVVGKIMAPEGISTLVLNIQRGILNVLQLNIKKTQNV
Liza_subviridis             AQLVTPIKFEYANGVVGKLLVPEGVSTMVLNIHRGILNLLQLNIKKTQNV
Melanogrammus_aeglefinus    SQLATPIKFEYANGVVGKVYSPEGISTTVLNIFRGVLNILQLNIKKTQNV
                            :** ******* ******:  ***:** ***: **:**:********:*:

Pagrus_major                YELQEAGAQGVCKTLYAIAEDEKAERILLTKTRDLNHCQEKIMKDLGLAY
Sillago_japonica            YELQEGGAQGVCKTLYAITEDEKADRILLTKTRDLNHCQEKIIKDIGLAY
Verasper_moseri             YELQEAGAQGVCKTLYAITEDEKAERILLTKSRDLNNCQEKIMKDIGLAY
Liza_subviridis             YELQEAGTQGVCKTLYAISEDERAERILLTKTRDMNQCQEKIIKDMGLAY
Melanogrammus_aeglefinus    YELQEAGAQGVCSTLYAITEDEKADRILLTKTRDLNNCQEKIVKDLGLAY
                            *****.*:****.*****:***:*:******:**:*:*****:**:****

Pagrus_major                TEKCAKCQQDSKNLRGATAYNYILKQAPSGIVILEAAVNELIQFSPFTEM
Sillago_japonica            IEKCPKCQQESKNLRGATAYNYILKPVASGILILEAAVNELIQFSPFNQM
Verasper_moseri             TEKCIKCQQ---NLRGAAAYNYILKPVANGILIQEATVNELIQFSPFDEM
Liza_subviridis             TEKCAKCQQDSKNLRGATSFNYILKPVASGVLILEVAVNEVIQFSPFAEL
Melanogrammus_aeglefinus    TEKCIECQKEAKNLRGAAAYNYVLKPVPTGVLILQATVNELIQFSPFNEL
                             *** :**:   *****:::**:** ...*::* :.:***:****** ::

Pagrus_major                NGAAQMQTKQSLVFLEIQKAPIVPITAQYLHRGSLKYEFSTELLQTPIQL
Sillago_japonica            NGATQMQTKQSLVFLEIQRAPIVPIEAQYLHRGSLKYEFSTELLQTPIQL
Verasper_moseri             NGAAQMETKQSLVFLEIQNAPIMPIKAEYLQRGSLKYEFSTELLQNPLQL
Liza_subviridis             KGAAQMETKQSLVFLEIQRAPIAPIEAQYIHQGSLKYEFSTELLQTPIQL
Melanogrammus_aeglefinus    NGAAQMNTKQSLVFLEIQKTPIAPVNAQYLHRGSIKYEFASELLQTPITL
                            :**:**:**********.:** *: *:*:::**:****::****.*: *

Pagrus_major                IKVNNVQTQIVEILNHLVTHNMQTVHEDAPLKFLELIQLLRSARFEDLEM
Sillago_japonica            VKVNNVQAQIVEVLNHLVSHNVERVHEDAPLKFLEFVQLLRAARLEDLEM
Verasper_moseri             IKIKNAQAQIVDILNHLVTHNVERVHEDAPLKFLELIQVLRAAQFEDLEM
Liza_subviridis             IKINNAQAQIVDVLNHLVIHNAERVHEDAPLKFLELIQLLRAARFEDLEM
Melanogrammus_aeglefinus    IKITNPQAQIVELLNHLVINNMQTVHEEAPIKFLELIQVLRKSNIADIDQ
                            :*:.* *:***::***** :* : ***:**:****::*:** :.: *::
```

```
Pagrus_major                LWSQYRTRPAYRQWILDAIPVIGTPAALKFIREKFLAHDLTVAEAAQALV
Sillago_japonica            IWSQYRARPASRQWILDAIPAVETPAALRFIKEKIMVDELTVAEAAQALI
Verasper_moseri             LWSQFRNRPAYRQWLLEAIPVIGTPAALKFIKEKYLADELTIAEAAQALI
Liza_subviridis             TWNIQKD-PCPQTVDPGCCPSHWKCCCSTIHHGEIPVRHN--CCSSSGLI
Melanogrammus_aeglefinus    IWSQFKGKPVYRQWILDTLPAIGTSTTLRFIREKLMARDISILEAAQALV
                             *.  :  *  :       *   .     : : :  . .     ::..*:

Pagrus_major                ASIHMVSASTEAIKQLEALAVNNKIVESPILREIILLGYGTMISKYCAEM
Sillago_japonica            ASIHMVTASPETIKLIEALSVNQKIVENPVLRQIVLLGYGTMISKYCVEM
Verasper_moseri             ASVHMVTASPEAIKLIEGLAINRKILEQPVLREIVLLGYGTMISKYCAEK
Liza_subviridis             ASVHMVTADTEAIKLIKSLAVNSKVIDNPVLREIVLLGYGTMISKHCVEL
Melanogrammus_aeglefinus    ASIHMVTADQEAITIVESLIDNNLIVENPALRQIVMLGYGTMVAKFCSDK
                            **:***:*. *:*. ::.*  *  :::.* **:*::******::*.* :

Pagrus_major                AVCPAELIKPIQDLLADAVAKADTQEIILLLKVLGNAGHPSSLKPITKIL
Sillago_japonica            TVCPAELIKPIQERLVEAVAQDDVQEIILLLKVLGNTGHPSSLKPITKLL
Verasper_moseri             TVCPAELVKPIQELLAEAVARAETQDTILLLKVLGNAGHTNSLKPITKML
Liza_subviridis             AVCPAELIKPIQDLLAEAVTKDETEDIILLLKVMGNAGHLASLKAITKIL
Melanogrammus_aeglefinus    IACPSELIKPIHEKIARAVAEENIEDIILLLKVLGNAGHPHSLKTITKIL
                             .**:**:***:: :. **:. : :: ******:**:**  ***.***:*

Pagrus_major                PIHGTAAASLPMKVHADAIMALRNIAKKEPRMIQELALQLYMDKALHPEL
Sillago_japonica            PIHGTPAASLPIRVHADAILALRNIAKKEPRMIQELTLQLYMDRALHPEL
Verasper_moseri             PIHGTAAASLPMRVHAEAIMALRNIAKKEPRMVQELALQLYMDKALHPEL
Liza_subviridis             PIHGTAAAVLPVRVHVQAIMALRNIAKKEPRMVQELALQLYMDKVLHPEL
Melanogrammus_aeglefinus    PLFGSAAAALPTRVHVDAILAMRNIAKKEPRLIQDMVVQLFMDKGLQPEL
                            *:.*:.** ** :**.:**:*:*********::*:::.:**:**: *:***

Pagrus_major                RMLACILLFETRPAMGLVTTLANIVKQEENLQVASFTYSHMKSLTRSTAA
Sillago_japonica            RMLSCIVLFETRPAMGLVTTLANIVKTEENLQVASFTYSHMKSLTRSTAA
Verasper_moseri             RMLACIVLFETRPPVGLVTTLASIVKAEENLQVASFTYSYMKSMTKSNSA
Liza_subviridis             RMIACIVLFETMPPMGLVTTVANIVKTEENLQVASFTYSHMKSLTRSSAA
Melanogrammus_aeglefinus    RMLACIVLFETKPRIGLVTALANVLKTEENLQVASFTYSHMKSLTRSTAS
                            **::**:**** * :****::*.::* ************:***:*:*.::
```

A

```
Pagrus_major                IHASVAAACNVAVKILSPKLNRLSMRFSKAIHVDSYYSPLMIGAAASAFY
Sillago_japonica            IHSSVAAACNVAVKILSPKLNRLNFRFSKAIHMDIYNNPLMLGAAASTFC
Verasper_moseri             IQESVSKACNVAVKILNPVLGRLSLRFSKAVYADVYNSPLMLGAAASAFY
Liza_subviridis             IHASVAAACNIAVKLLSPRLDRLSLRFSKAIHVDVYNNPLMLGAAAGAFY
Melanogrammus_aeglefinus    HQASVAAACNVAVRILSPKLDRLSYRFSKAIQMDTYSEPMMLGAAASAFY
                             : **: ***:**::*.* *.**. *****:  * * .*:*:****.:*

Pagrus_major                INDAATILPRSFVAKTSAYLAGAAADVLEVGVRTEGLQEALLKNPALINN
Sillago_japonica            INDAATVLPRSIVAKTSAYLAGAAADALEVGVRAEGHQEAFLKNPALIEN
Verasper_moseri             INDAATLLPRSFVAKAGAFLAGAAADVVEVGVRTEGLQEALLKNPAVVDS
Liza_subviridis             INDAAT-------------------NVLEVGVRTEGLQEALLKNPVLIDN
Melanogrammus_aeglefinus    INDAASILPRAFVAKTRGYLAGAAADVLEIGVRTEGLQEALLKNPD-NEN
                            *****:                      :.:*:***:** ***:****   :.

Pagrus_major                ADRMTKMKRVIKALSELRSLPARTPLASVYIKFFGQEIAFANIDKNIIDQ
Sillago_japonica            ADRITKMKRVIKALSDLRSLPISKPLASIYVKIFGQEIAFANINRALIEQ
Verasper_moseri             ADRITKMKRIMQALSHWRSMPNRSPLASVYVKFFGQEIAFANIDKALIDQ
Liza_subviridis             ADRITKMKRVIKALSHWRSLPSTQPLASVYIKFFGQEIAFANIDKALIDQ
Melanogrammus_aeglefinus    ADRITKMRRVIKALSQWRSLPTSKPLANVYIKFLGQEIAFANLDKVLIDQ
                            ***:***:*:::***. **:*   ***.:*:*::********::: :*:*

Pagrus_major                AIALATGPSLHTFGKNAIKALLTGASFHVAKPLLATEVRRIMPTAAGLPM
Sillago_japonica            AIALATGPSAQQFGKNAIRALLSGTSFHVAKPLLATEMRRILPTAAGFPM
Verasper_moseri             VISLATAPSAQALGRNAVKALLSGASFNFAKPLLATEVRRIMPTAAGLPM
Liza_subviridis             AIALATGPSVQALGRNAIKTLLSGASLHFAKPLLITEMRRILPTAAGLPM
Melanogrammus_aeglefinus    AIALASRPIVQEMGKDALRSVLSGYTAHFTKPVLAAEVRRILPTAAGFPM
                            .*:**: *  : :*::*::::*:* : :.:**:* :*:***:*****:**

Pagrus_major                ELSLYTAAVAAAAVQVKATTTPALPENFLLTHLLKTDIQFETEIRPSVAV
Sillago_japonica            ELSLYTVAVAAAGVQVKATSTPALPENFLMAHLLKTDMQLQTEIRPSIAM
Verasper_moseri             ELSLYTAAVAAVAVQIKAITSPVLPENFHLAHLLKTDMQLETEIRPSVAV
Liza_subviridis             ELSLYTAAVAAAGVKVQAKTVPALPENFHFAQLLKTDIQLETEIRPSVAM
Melanogrammus_aeglefinus    ELSMYTAAVANAAVEIKASATPALPQNFLWTHLLKANINLETEIRPSVAV
                            ***:**.*** ..*:::* : *.**:**  ::***::::::******:*:

Pagrus_major                NTFAVMGVNTAMLQAALISRAKLNSIVPAKIAARLDINEGHFKIEALPVS
Sillago_japonica            NTFAAMGVNTAILQAVLLSRAKLNSIVPAKIAARLDINEGNFKIEALPVP
Verasper_moseri             NTFAVMGVNTAILQAGLQSRAKLNSILPAKIIATLNIQEGHFKLEVLPVS
Liza_subviridis             NTFAVMGVNTAVLQAAVLSKAKLNSLVPAKIAARLDINEGHFKIEALPGT
Melanogrammus_aeglefinus    QSFAVMGVNTDLIQSSVLVRAKVNSVLPSKIAANLDITGGNFRIQALPVV
                            ::**.***** ::*: :  :**:**::*:** * *:*  *:*:::.**

Pagrus_major                VPENIATVHVETFAVARNIEDLAAARITPIIPAKVLKPFSREILTSKLAS
Sillago_japonica            VPEEVAAVHVETFAVARNIEDLAAPKLIPIIPAKVLQPLSREILSSKIVS
Verasper_moseri             APENIAAVQVDTFAVARNIEDLAAARIVPLIPAKLMQPISR---LSKHSS
Liza_subviridis             VPENTAVVHVETPAVAR--------------------------------
Melanogrammus_aeglefinus    VPEQLAAVRYESFAVARNIEDLLAEKIVPIIPAEILEPLIGRRFVSNANS
                            .**: *.*: :: ****
```

B

```
Pagrus_major                NKFLGNEAPPAFAIIVRAVRADNKMMGYQLAVYLDKPSTRIQIILAALAA
Sillago_japonica            NKFLGNAVAPTFAIIARAVKADKRVLGYELAVYLDKPTARLQIILANLAA
Verasper_moseri             NKYLGKEATPVFAIILRAIRADKKMMGYELSVYLDKPTARIQLILAALAA
Liza_subviridis             NKFLGNDAAPTFAIIFRAIRADNMMQGYQLAAYLDKPSARLQIILAALAL
Melanogrammus_aeglefinus    NKYLGASAPPAFAILVRVVRANNQMMGYELSTYLDKPNKRVQLIIAALAK
                            **:**  ..*.***: *.::*:: : **:*:.*****. *:*:*:* **

Pagrus_major                DNNWKLCADGALLSKHKVTAKIGWGAECKQYDTMITAETGLVGPSPAARV
Sillago_japonica            DNNWKICADGALLSKHKVTAKIGWGAECKQYDTIITAETGLVGPSPAARL
Verasper_moseri             DDNWKLCADGVLLSKFKVNAKVGWGAECKQYDTMITAETGLVGPRPAARI
Liza_subviridis             T-------------------------------------------------
Melanogrammus_aeglefinus    GDNWKLCADGILLSKHKVTAKIAWGAECKEYSTMATAETGLVVKSPAARL

Pagrus_major                RVAWNDLPSAIKHYAKKMYDLIPANMLPGLIKGKDENSANQLSMTVIATS
Sillago_japonica            RVAWNELPSAFKHYARKVYDYIPASSVPGLIKGKDETSAKQLSLTVVATS
Verasper_moseri             RVAWNNLPTALKRYAEKVYNAIPASMLAGLIQAKDEKSSNQLSATLIATS
Liza_subviridis             -------------------------------------------------
Melanogrammus_aeglefinus    TVSWETLPISFKTYAKLIYKYIPASILAGLVEGKEINIEKQVSLIVVATS

Pagrus_major                DRTIDFIWKSPTRTFYKLALHLPYPLPLDGIKGLTPFDG-LADQVHYLFA
Sillago_japonica            DKTFDFVWKIPTNTVYKLAVHLPIALPLDEFKGLTPFDE-LADQIHYVFA
Verasper_moseri             DRELDLVWKTPIR-VYKLSLRLPIALPLDAIKGLTPFDD-AADRVHYWLA
Liza_subviridis             -------------------------------------------------
Melanogrammus_aeglefinus    NKDLDLIMKTPKHTIYKLALHLPIALPLDGLTNLTPFEDNIMEKVRYVLA

Pagrus_major                KAAAAECSFNRDALTTFNGRKYKNEMPLSCYQVLAQDCTNELKFMVLLKK
Sillago_japonica            KAIAAECSFRENTLTTFNSRRYRNEMPLSCYQVLAQDCTDELKFMVLLKR
Verasper_moseri             KAGAAECSFARDTLTTFNNRKYRTEFPLSCYQVLAQDCTDELKFMVLLKK
Liza_subviridis             ----------------------------LSCNQILAQDCTDEMKFMVLLKK
Melanogrammus_aeglefinus    QVYEAQCSYAKNTLTTFNNRRYKNEMPLSCNQVLAQDCSKELKFMVLLKK
                                                        *** *:*****:.*:*******:

Pagrus_major                DHIEQNHINVKIADIDIDLYPKNTDVIVKVNGMEIPINNLPYQHPTAKIQ
Sillago_japonica            DHIEQSHINVKIADIDIDLYPRNTDVIVKVNGMEIPINNLPYQHPTAKIQ
Verasper_moseri             DNIEQNHINVKIADIDIDLYPKNADVILKVNGMEIPISNLPYQHPTAKIQ
Liza_subviridis             DHIKQNHINVKIADIDIDLYPKNSDVVVKVNGMEMPISNLPYQHPTANIQ
Melanogrammus_aeglefinus    DNIEQNWINVKLADIDIDLYPKNNDVIVKVNGMQIPISNLPYQHPTGTIQ
                            *:*:*. ****:*********:* **::*****::**.********..**

Pagrus_major                IRPKGEGISVYAPSLGLHEVYFDRNSWTVKVVDWMKGQTCGLCGKADGEV
Sillago_japonica            IRPMGEGISVFAPSHGLQEVNFDRNSWKVKVADWMKGQTCGLCGKADGEV
Verasper_moseri             IRPKEEGISVFAPSLGLHEVYFDKNSWKVKVVDWMKGQTCGLCGKADGEV
Liza_subviridis             IRSKGEGIAMYASSHGLHEVYFDKNSWKIKVVDWMKGKTCGLCGKADGEI
Melanogrammus_aeglefinus    IKSNGEGISVYAASHGLHEVYFDKSSWKVKLADWMKGQTCGICGKADGEV
                            *:.  ***:::*.* **:** **:.**.:*:.*****:***:*******:

Pagrus_major                RQEYRTPNGRVTKSAVSYAHSWVLPAESCRDTTECRMKLESVQLEKQVNI
Sillago_japonica            KQEYSSPSGRVTKSAVSYAHSWVLPAESCRDTTECRMKLESVRLERQITV
Verasper_moseri             QQEYHTPNGRLTKNAVSYAHSWVLPAESCRDTTECRMKLESVQLQKQLNI
Liza_subviridis             RQEYRTPNGRLTKNAVSFAHSWILPAESCRDTTGCRVKLESVQLEKQVNV
Melanogrammus_aeglefinus    RQEFRTPNGQLAKDAVSYAHSWVIPAENCQDASECRMRQESVKLERQIII
                            :**: :*.*:::*.***:****::***.*:*:: **:: ***:*::*: :
```

C

Figure 4.45 Alignment of lipoprotein N-terminal domain (A), lipovitellin-phosvitin complex; beta-sheet shell regions (B), and von Willebrand factor (vWF) type D domain (C) of fish Vtg-1. Blue label is lipoprotein N-terminal domain, green label is lipovitellin-phosvitin complex; beta-sheet shell regions, and yellow label is von Willebrand factor (vWF) type D domain.

Additionally genomic sequences at 3'side of *vtg*-1 were determined by GenomeWalk PCR. The results of PCR product was shown in Fig. 4.46.

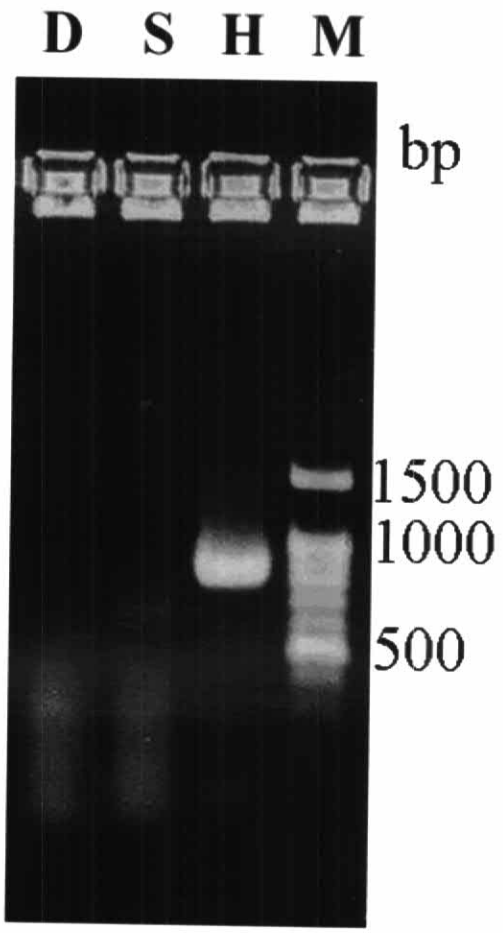


Figure 4.46 3' GenomeWalk PCR product of *vtg*-1. Lane M, D, S and H are 100 bp markers, 3' GenomeWalk PCR product of *vtg*-1/*Dra*I, *Ssp*I and *Hae*III, respectively.

3'GenomeWalk PCR product of *vtg*-1/*Hae*III was cloned and sequenced (2 different size of insert) as shown in appendix B and BLAST analysis found it homologous with *vtg* of orange-spotted grouper (*Epinephelus coioides*) with E-value as 5e-40 and 1e-64 , respectively (see in appendix C).

4.5.2 *vtg*-3

Partial cDNA sequence of *vtg*-3 at the sizes of 500 bp and 1 kb was amplified from 1st strand cDNA templates of liver (Fig. 4.47 A and B, respectively).

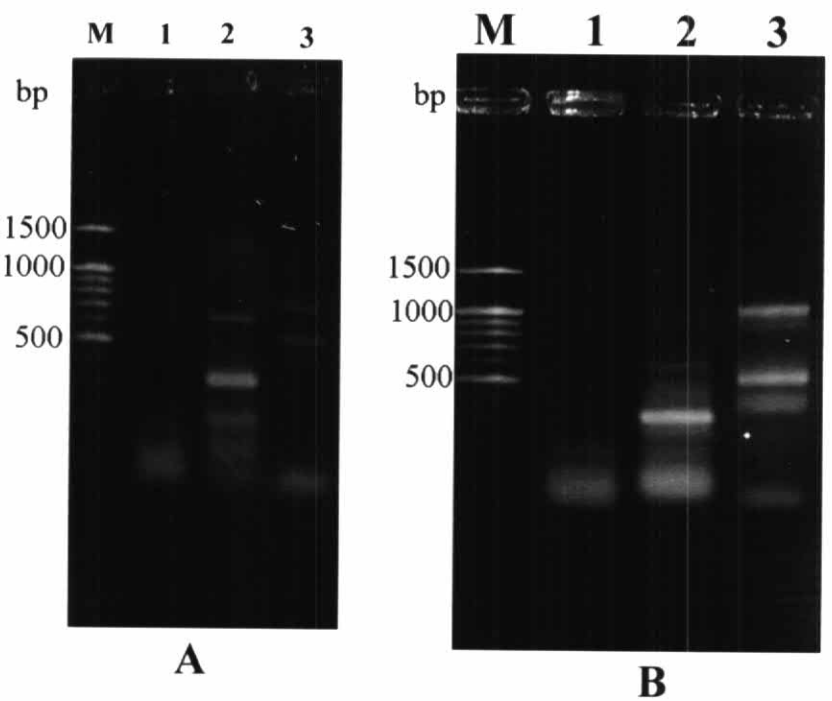


Figure 4.47 PCR product of *vtg*-3 at the sizes of 500 bp (A) and 1 kb (B). Lane M is 100 bp markers, lane 1, 2, 3A, 3B are negative control, positive control, PCR product of *vtg*-3 at size 500 bp and 1 kb, respectively.

The results of sequencing and BLAST analysis indicated that both sequences were homologous with phosvitinless vitellogenin of red seabream (*Pagrus major*) with E-value as 9e-82 and 4e-120, respectively (see in appendix B and C). Resulting band from 5' RACE PCR amplification at the size of 1.5 kb (Fig. 4.48) was cloned and sequenced. BLAST analysis indicates that it was homologous with phosvitinless vitellogenin of red seabream (*Pagrus major*) with E-value as 5e-98. New RACE primer was designed from this sequence and used in 5' RACE PCR amplification. The result in Fig. 4.49 showed bands at the sizes of 1.8, 1.4 and 0.9 kb. The sequencing and BLAST analysis indicated that these bands were homologous with phosvitinless vitellogenin of red seabream (*Pagrus major*) with E-value as 0.0, 1e-178, and 1e-92, respectively (see in appendix B and C).

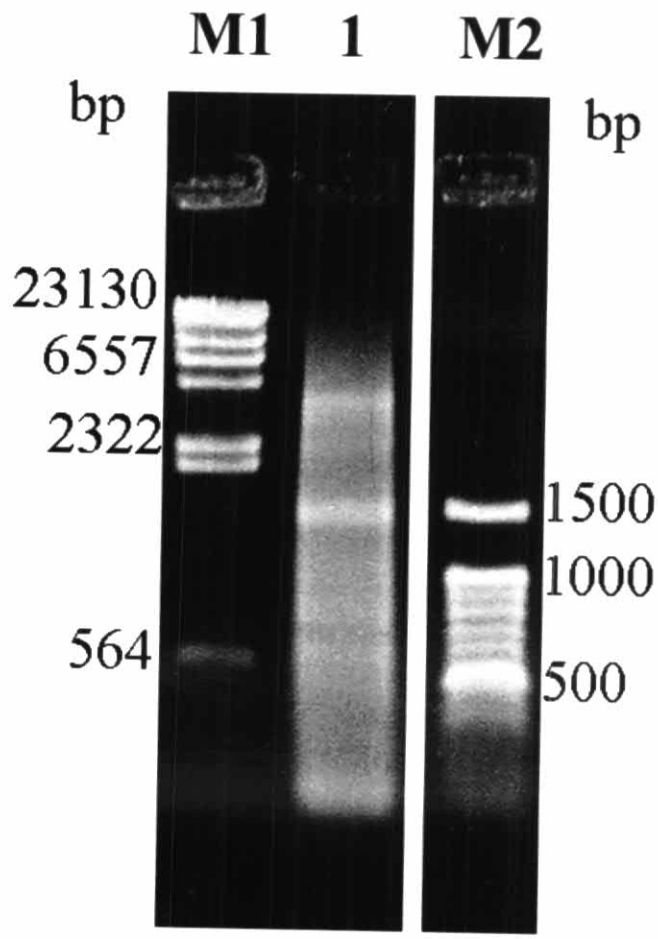


Figure 4.48 5' RACE PCR product of *vtg*-3 separated in 1.2 % agarose gel. Lane M1, M2 and 1 is λ*Hin*dIII, 100 bp DNA marker and 5' RACE PCR product of *vtg*-3, respectively.

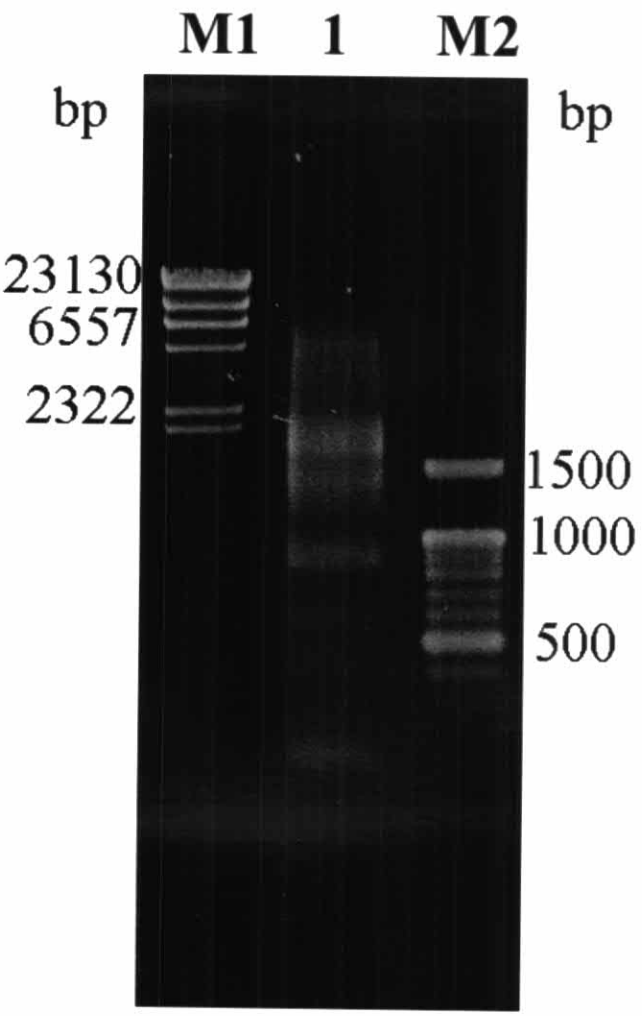


Figure 4.49 5' RACE PCR product of *vtg*-3 (continued) separated in 1.2 % agarose gel. Lane M1, M2 and 1 is λ*Hin*dIII, 100 bp DNA marker and 5'RACE PCR product of *vtg*-3 (continued), respectively.

Nucleotide sequences of *vtg*-3 (*in silico* ligated) divided to 2 parts (5' and 3' side of cDNA sequence) was translated to amino acid sequence of N and C-terminus of Vtg-3, respectively (Fig. 4.50). *L. subviridis* Vtg-3 contained lipoprotein N-terminal domain (LPD_N) and lipovitellin-phosvitin complex; beta-sheet shell regions. LPD_N and beta-sheet shell regions of *L. subviridis* Vtg-3 were aligned with these domains of other fish species (Fig. 4.51 (A), (B), and (C), respectively). LPD_N and beta-sheet shell regions of *L. subviridis* Vtg-3 were conserved with these domains of other fish species.

```
RGVSDLAFQEFNGFPGKNGFNASPKLTKRIAAQLTKPFMFEFASGHVGDIRAAAEISDTVVNIVRGILGFFQVTVKTTQRIYEL
EEVGIHGKCQSNYAMEENTETQDMTITQVVDVSNCREKAAIYRGMATAVLDRVSKQRGESVISTVRYVYTVKPTAEGGLITRAH
GLERQHFSPLNVKGGSFKMQAMKEMVLLGVSDTARAITFGPMESKGNLVYKFVRAANIPIIMENLDNPIPKAVELIKQLAQANR
YQVDSATTEDTIKLYQLLRVMPYEGLDVMWAQLPGNEEHRRWFLDMIVEIGDARILKFLETRFKAGDVSASEALETLLLSINHL
QAIPELVEMAKVFLTMPFSKSNTFLWHTVVLAYGSLVYKHCAYYTPCPVSAVQPLLDVATESLRNNNEEDMVLALKALGNAGHP
GSIKTIMRFLPGVAANPVDLPPRVLSAAVQSMRLTAARDRHSVQGITLSLFLQKHLPAEIRMMAITILFDTNPSMALVSTVTIH
LLEEKDFHVVSFVYSYLQSVARSTTPDNHFLSTACNVAVKILAPKFGRTSYYYSQARRMDWFSDDFLIGTAAEVFMLRSANNIF
PTEFITKGKFYFIGRILQLLELGIRAEGLKELFGASIPGFKGDFSFSDFRAIFNVLQNWEILPNNKPVLSAYLRASGQECFFGD
INKYLIQNIVRAFSPSAGKESPVFAAIQNLQKGISWHRTKPFLIFEARYFQATTLGLPVEISKYYETVYGITVNAKAAVNPPPT
EHLAQLLNSEISLESDGFVGFTKGFLGFYGTNTELFPSGLQIEKQNASLP

                                                A

HPGKFASQVQVGKKIGEGFAGFKKGEVFGTDLEDFGQVQRCGVYWKANSFYSNFFIQGNLPCRFFEGLGSLGVQKAIVWQPRFE
MCVGIKFRVGFWVEYEAWRLDSEEYPLNVLLGFTYIAVKLLRSRQEKLVTKFTLKFMPTPADIQCVHAQDLRFCGRFPKKQPGE
QSQSSDSASSDRGSHQSKRDMIMGGWESTLEDVFNIKTFDQVVNQKPEGYDASVYYTPEANTQNAQLIVSQVGEDTNWKMCVDT
TVNAGSGAKAHIRWGAECQSYEISMRAATAYLPGSKPALKAKVHWTRVPEAMEDMGTRIESYIPGMAFLFGFYQQNERNAAQEV
SALVVAESADSVDVKIKFPKFTVYHQAIPFLLPPANFQEFHPSITNTTVDAGRA

                                                B
```

Figure 4.50 N-terminus amino acid sequences (A) and C-terminus amino acid sequences of *L. subviridis* Vtg-3 (B). Yellow label is lipoprotein N-terminal domain and blue label is lipovitellin-phosvitin complex; beta-sheet shell regions.

```
Liza_subviridis              --------------------RGVSDLAFQEFNGFPGKNGFNASPKLTKRI
Pagrus_major                 SGVRMTCRVKIVGASAQTFVLQVSDLAFEEFNGFPGKNGFNASPKLTQRI
Gambusia_affinis             SGVRITCKVSIIGIDKQSFLLQVSELEFAEYNGFEGKDDYVSSPKVTKRI
Acanthogobius_flavimanus     SGARLKTIVTIMGAQSDTFHLQFSNVVFEEFNGVPGKSDFVAVNQLAQKI
Danio_rerio                  SALKLRCTFKIIGESPHTFVLQVSNVDFEDFNGIPGKSVFSPSKNITKYL
                                                .*:: * ::**. **. : .  :::: :
Liza_subviridis              AAQLTKPFMFEFASGHVGDIRAAAEISDTVVNIVRGILGFFQVTVKTTQR
Pagrus_major                 AAQIVKPFMFDYTSGHVGDIRAPAEVSDTVVNIVRGILGFLQVTVKTTQT
Gambusia_affinis             TSELAQPFLFEYDNGHIGDISGHSKVSTTVVNIIRGILSFFQVTVKNTQT
Acanthogobius_flavimanus     AAQLEKPIGFTYTAGNVGEIKASPEVSDTVVNIVRGILGFFQVTVKSNQD
Danio_rerio                  SAEISQPIISEYSKGQITDIRTAPGVSNTVVNIVRGILGFLQVTVKTTQS
                             :::: :*:   :  *:: :*   . :* *****:****.*:*****..*
Liza_subviridis              IYELEEVGIHGKCQSNYAMEENTETQDMTITQVVDVSNCREKAAIYRGMA
Pagrus_major                 VYELEEVGIHGKCQSNYATEENTETKDMTITQVVDVGACKEKAAIYRGMA
Gambusia_affinis             IYELRETGIHGNCHNDYAIEEKEGTKDWIVTQVVDVTNCWERAAMYSGMA
Acanthogobius_flavimanus     IYELEEIGIHGKCLSSYATKINEQEKVMDLTQVVDVTNCREKAMFQTGMA
Danio_rerio                  FYELIELGIHGLCQSSYTVDEDSNPKELIVTRIVDITNCQQPASLYRGMA
                             .*** * **** * ..*: . .   :   :*::**:  * : * :  ***
Liza_subviridis              TAVLDRVSKQRGESVISTVRYVYTVKPTAEGGLITRAHGLERQHFSPLNV
Pagrus_major                 TAVLDQVSKQRGESVISTVRYVYTVKPTAEGGLITRAHGLERQHFSPFNV
Gambusia_affinis             TAEWDRLAKERGENVISSMKYNYTIKPTGDGGLITRAQGFERQHFSAFNV
Acanthogobius_flavimanus     TAVEDKVSRQRGESVFSTVKYTYNIKATEEAGLITKAQALELQFFTPFNL
Danio_rerio                  LAPEDKLSKQRGESVVSTVKHTYTVKSTADGGQITKAFAQERQYFSPFNV
                              *  *:::::***.*.*:::: *.:*.* :.* **:* . * *.*:.:*:
```

```
Liza_subviridis                KGGSFKMQAMKEMVLLGVSDTARAITFGPMESKGNLVYKFVR-AANIPII
Pagrus_major                   KGGSFKMQAMKEMVLLCVSDTARAFTYGPMESKGNLVYKFVNAEANVPIM
Gambusia_affinis               KGGSFKMEATKDLVLLSMNRTARGRTYGPLEKKGKIIYSFEDVDINIPTM
Acanthogobius_flavimanus       KGGTFKMEAMKELVLTTVKDKTQDTPNRQMESRGNIVYKVVKNWANVPIM
Danio_rerio                    KGGNFRMLALRDIELLKVSDTTDKVVTGQVQSRGNLMYKTNKDLKPIPVV
                               ***.*:* * ::: *  :. .:       ::.:*:::*.       :* :

Liza_subviridis                MENLDNPIPKAVELIKQLAQANRYQVDSATTEDTIKLYQLLRVMPYEGLD
Pagrus_major                   MQNLDDPVPKAIELIKQLAEANKYEVDSATTEVTIQLYQLLRVIPYEGLD
Gambusia_affinis               MQKLDNPGPKATELIKRLSEANSGTINSATTEDSIKLYHLLRVTPYEELE
Acanthogobius_flavimanus       MQKLDDPVPKATELIKRLVQANTNQLDSTTTEDAIKLYQLLRVIPLEKLE
Danio_rerio                    MLNLNDPVPKILDLIKRLAQANIYHVDSETSTEILDLIQLMRVTTLDNLE
                               * :*::* **  :***:* :**   ::* *:   :.* :*:** . : *:

Liza_subviridis                VMWAQLPGNEEHRRWFLDMIVEIGDARILKFLETRFKAGDVSASEALETL
Pagrus_major                   TMWKQFAGNEQYRHWFLDMIVEVSDARILKFLETRFQAGDVSPSEALQTV
Gambusia_affinis               NMWKQLAVNPKYRRWFLDSIVEIADIKVLNFMEARFKANDWTHFEALQTI
Acanthogobius_flavimanus       KMWMEVEHNRNERNWFLQTVVEVNDARILALLERLLREHKLNRLEAITVI
Danio_rerio                    HLWKQVSGNDEHRRWFLDLVVEVTDERILKFLEARYKAGDITANEAGQAL
                                :* :.  * : *.***: :**: * ::* ::*   :  . .  **  .:

Liza_subviridis                LLSINHLQAIPELVEMAKVFLTMPFSKSNTFLWHTVVLAYGSLVYKHCAY
Pagrus_major                   LLSINHLQPIPELVEMAKMFLNMPFSKSNTYLWHTVVLTYGSLVYKHCAY
Gambusia_affinis               LIAFHHLQPTPENVGMAKIFLKLPFSKSNTFLWHTVVLSYGSLVNKYCVH
Acanthogobius_flavimanus       VRAFNHLEATPELLREAEKFLTMTYN--DAMIRRTVVLSFGSLVYRHCAY
Danio_rerio                    VVAFNHLSAEPVSVALAQESLTIPFSKSHPLLWNTVVLAYGSLVHRYCVY
                               : :::**.. *  :  *:  *.:.:.  .. : .****::**** ::*.:

Liza_subviridis                YTPCPVSAVQPLLDVATESLRNNNEEDMVLALKALGNAGHPGSIKTIMRF
Pagrus_major                   YTPCPVNAVQPLLDMALESLRNGNEADMVLALKALGNAGHPGSIKTIMRF
Gambusia_affinis               YYICSVDVIQPLVEMATEALRNGNKEEMIIALKAMGNAGHPKSMKTIMRF
Acanthogobius_flavimanus       NTPCPEEAIRPLLNMAEESRRNNNEMEQILVLKALGNAGHPRSLKAIEKF
Danio_rerio                    TDPCPITVVQPLLNMAASSLSKNSEDEMVLALKSLGNAAHLSSIKTLLKF
                                  *.  .::**:::* .:  :..: : ::.**::***.*  *:*:: :*

Liza_subviridis                LPGVAANP--VDLPPRVLSAAVQSMRLTAARDRHSVQGITLSLFLQKHLP
Pagrus_major                   LPGVAATP--VDLPPSVLSAAVQSMRLIAARDPHSVQDITMSLFLQKNLP
Gambusia_affinis               LPGVSATP--VDLPLRVQSAAIQAMRLVANRDPHNVREAALAVFLHKHVA
Acanthogobius_flavimanus       LPGVDPNTNQADSKPRVVSAAVQAMRLIATREPHGVQLRTLKLFLNRELK
Danio_rerio                    LPGYSNGA--EKLSTRVQGAAVQAFRLLASRAPHSVQDIVLNLFVQKHLP
                               ***    .   .    * .**:*::** * *  *.*:  .: :*::::.:

Liza_subviridis                AEIRMMAITILFDTNPSMALVSTVTIHLLEEKDFHVVSFVYSYLQSVARS
Pagrus_major                   TELRMLSFMILFDTKPPMALVSTVTAHLMEEKDLHVVSFAYSYLRSLGRS
Gambusia_affinis               SEVRMHAFKILFATKPSMALVSTVTAHLQDEKDLQVASFAYTYLHGLARS
Acanthogobius_flavimanus       PEIRMLALMIMFDTKPSIGLVSTVTAHLLEERDMQVVNFAYTYFKSLSRS
Danio_rerio                    AEIRMLACIVLLETMPSTALISVVSEVLLEEADLQVASFSYSLLKGFAKS
                               .*:** :  ::: * *. .*:*.*:  * :* *::*..* *: ::...:*

                                                    A

Liza_subviridis                TTPDNHFLSTACNVAVKILAPKFGRTSYYYSQARRMDWFSDDFLIGTAAE
Pagrus_major                   STPENFFLSTACNVAAKILAPKFGRLSYHYSKAMRMDWFNEDFLIGTAAE
Gambusia_affinis               RTPNNQYRSIASSLAVKILAPRFGRLCYYYSKTKHADWFDDRWLTGMTSE
Acanthogobius_flavimanus       MTPDNHFLSTAASVAVKILAPKFCHLSYYRSRATRMDWFSDDYLIGTAAE
Danio_rerio                    RTPDNQHLSIACNIAMKILTRKLGHLSYRYSKNLHFDWFHDDFLFGTSAD
                                **:* . * *..:* ***: :: : .*  *:  : *** : :* * :::

Liza_subviridis                VFMLRSANNIFPTEFITKGKFYFIGRILQLLELGIRAEGLKELFGASIPG
Pagrus_major                   VFMLRSATNIFPTEIMMKGKFFFIGRILQLLEFGIRADGLKDLFGT---G
Gambusia_affinis               FFMLRNASNAIPSEIVSSGKLHIIGRIIELLELGVRADGIKDLIGAGIPG
Acanthogobius_flavimanus       AFILKRASQVIPAGLMVKWNFHFIGRILQLVEFGIRPEGLRDLFGASVPE
Danio_rerio                    VYMLQNES-PIPTKLMLKGKFHFIGRILQFLEFGIRADGLKDLFAGKIPE
                                ::*:  .  :*: :: . ::.:****::::*:*:*.:*:::*:.

Liza_subviridis                FKGDFSFSDFRAIFNVLQNWEILPNNKPVLSAYLRASGQECFFGDINKYL
Pagrus_major                   FNGDLSFSDFQSLFNVLQNWETMPNNKPILSAFSRASGQEWFFADVNKEF
Gambusia_affinis               FKGDFSMSDLMAIFSVLQRWESLDNDKPVLSVYSRITGQEWFFADMSKET
Acanthogobius_flavimanus       FRGDMSMSDFQAIYDVLKKWETLPDDKPLLSVYTRASGQEFFFNDFNKDF
Danio_rerio                    LTKDLGISDLASILKILSNWQSLPKDKPLLTAYARVFGQEAFLMDVSRDS
                               :  *:.:**: :: .:*..*: : .:**:*:.: *  *** *: *..:

Liza_subviridis                IQNIVRAFSPSAGKESPVFAAIQNLQKGISWHRTKPFLIFEARYFQATTL
Pagrus_major                   IQNIIGTVSLSAGRESPLWTAIDNLQRGVSWHRVKPFLIFEVRYFQATTL
Gambusia_affinis               LRNIITTFSPAAWKSSPLKAMIDSLQKGISWDYTRTFLILEARYFQATTV
Acanthogobius_flavimanus       MERMMREFSPTAGRDSFVWRMIEQLIHGFSWRGVLPFLTVEARYIQATTL
Danio_rerio                    VQSIIKSFSPSAGKESKVWERIQDVQKGTSWHWTKPHLVYEARFIQPTCL
                               :. ::  .* :* :.* :   *:.: :* **  . ..*  *.*::*.* :

Liza_subviridis                GLPVEISKYYETVYGITVNAKAAVNPPPTEHLAQLLNSEISLESDGFVGF
Pagrus_major                   GLPLEISKYYESVNGITVNAKAAVNPPLTERLGQLLNSEISLESDGFIGY
Gambusia_affinis               GLPVEISKYYNAITAVNFNAKVAINPPPTESIGQLLNSEITMETDGYAGC
Acanthogobius_flavimanus       GLPVEISKYYHTVNAMSVNAKAAISPQMTDNVGQLLDAETTLKTDGFIGY
Danio_rerio                    GLPVEISKYYSVVNAVTMQAKAEINPPPKEHLGELLSSDISMQTDGFIGV
                               ***:******  : .:...:**. :.*  .: :.:**.:: ::::**: *
```

Figure 4.51 Alignment of lipoprotein N-terminal domain (A), lipovitellin-phosvitin complex; beta-sheet shell regions on N-terminus (B), and C-terminus (C) of fish Vtg-3. Yellow label is lipoprotein N-terminal domain, blue label is lipovitellin-phosvitin complex; beta-sheet shell regions.

4.6 ERE determination

ERE of *chg*-L gene was determined by amplifying 5' end of gene using genome walking technique. Genome Walk R primer designed from the DNA fragment of *chg*-L 900 and used in GenomeWalk PCR amplification with *Dra*I, *Ssp*I, and *Hae*III GenomeWalk library.

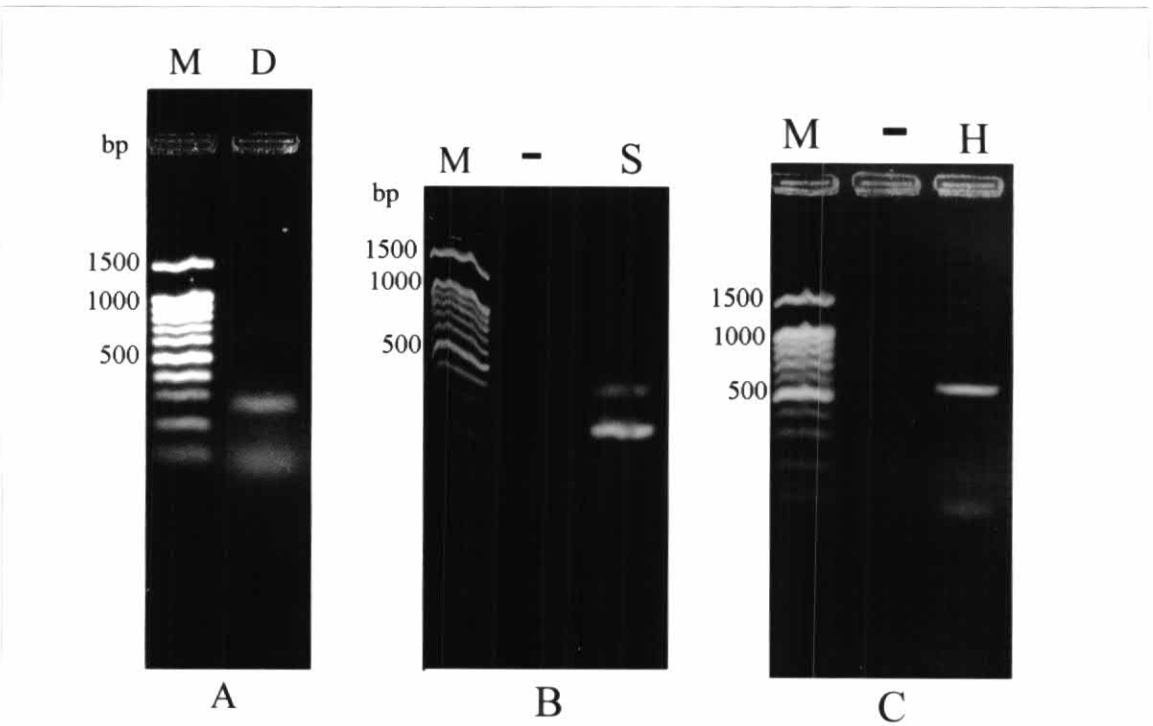


Figure 4.52 Secondary 5' GenomeWalk PCR product of *chg*-L which amplified from *Dra*I (A), *Ssp*I (B), and *Hae*III (C) GenomeWalk library. The products were separated in 1.5 % agarose gel. Lane M, -, D, S, and H are 100 bp DNA marker, negative control, *chg*-L/*Dra*I, *Ssp*I, and *Hae*III GenomeWalk product, respectively.

The result of secondary PCR product was shown in Fig. 4.52. The primary PCR products were smear. After secondary PCR amplification, the products at the sizes of 650 (Fig. 4.52 A) and 350 bp (Fig. 4.52 B) were cloned, sequenced and BLAST analysis (see in appendix C) found it homologous with chorion protein (*Liparis atlanticus*) and choriogenin L (*Oryzias latipes*) with E-value as 6e-33 and 3e-12, respectively. *chg*-L/D12 sequence which shown exons and introns position shown in Fig. 4.53. Exons were join and translated to amino acid sequence as shown in Fig. 4.54.

CGACGGCCTGGGCAGGTGTTGCATCTGTTGCTGTTGAGTGCAGAGAGAATGATGCT
CATGTGGAAGTCAGGAGGGATATGTTTGGGACTGGCCAGTTGGTCAATCCGAATGA
CCTCACCCTGGGGAACTGTCCTGCTGTCGCAGAGGATCCTGCGGCTCAAGTGTTGA
TTTTTGAAGCTGAACTGCATGACTGTTTGAGCTCATTGGTGGTAAGTAAAATACAT
AATGCTGAAACTAAACAATATCTTATAATGAATCGTATATGAATATATCATCAAGG
TTTTTGGTCATCTTTGCTAGTTGACAGAAGATTCCCTGACCTACATCTTCACTCTG
AACTACGATCCCCGACCTCTGGGTTCCTCCCCCGTAGTAAGGACCGGCAGTGCAGC
TGTTATTGTGGAATGCCACTACCCAAGGTGTGTTGACTGAAGATATGTGGCAATAA
ACAATCAAAACACAAGTAAAATAATCATCTTAATATGTAACGTAATGTGTAGCCTT
TAGCAGAAATGCACCGTGAAATCTAATCTGCACATGCCAGTGGTAGACGCATGTCA
CGCTGAGGATTTCTTTAGTCTTCTGTTTTAAGCGGGTATGTCTGTCTTGTCTCTTA
GAAAGCACAATGTGAGCAGCCTTCCTC

Figure 4.53 Nucleotide sequence of *chg*-L/D12. Blue label is exon and non label is introns.

```
gttgcatctgttgctgttgagtgcagagagaatgatgctcatgtggaagtcaggagggat
 V  A  S  V  A  V  E  C  R  E  N  D  A  H  V  E  V  R  R  D
atgtttgggactggccagttggtcaatccgaatgacctcaccctggggaactgtcctgct
 M  F  G  T  G  Q  L  V  N  P  N  D  L  T  L  G  N  C  P  A
gtcgcagaggatcctgcggctcaagtgttgatttttgaagctgaactgcatgactgtttg
 V  A  E  D  P  A  A  Q  V  L  I  F  E  A  E  L  H  D  C  L
agctcattggtggtaagtttgacagaagattccctgacctacatcttcactctgaactac
 S  S  L  V  V  S  L  T  E  D  S  L  T  Y  I  F  T  L  N  Y
gatccccgacctctgggttcctcccccgtagtaaggaccggcagtgcagctgttattgtg
 D  P  R  P  L  G  S  S  P  V  V  R  T  G  S  A  A  V  I  V
gaatgccactacccaaggaagcacaatgtgagcagccttcctc
 E  C  H  Y  P  R  K  H  N  V  S  S  L  P
```

Figure 4.54 Amino acid sequence translated from *chg*-L/D12 sequence.

chg-L/D12 GW primer were designed from this sequence and used for amplified further genomic sequence with GenomeWalk PCR amplification. *chg*-L GenomeWalk PCR product at size 900 bp was amplified from GenomeWalk *Ssp*I library (Fig. 4.55). After sequenced (see in appendix B) and BLAST analysis found it homologous with chorion protein (*Sparus aurata*) with E-value as 8e-14. Additionally GenomeWalk PCR amplification of *chg*-L was do again (used EXT *Taq* in amplification), DNA fragment at size 1.76 kb and 900 bp were amplified from *Dra*I and *Ssp*I GenomeWalk library, respectively (result shown in Fig. 4.56 A)

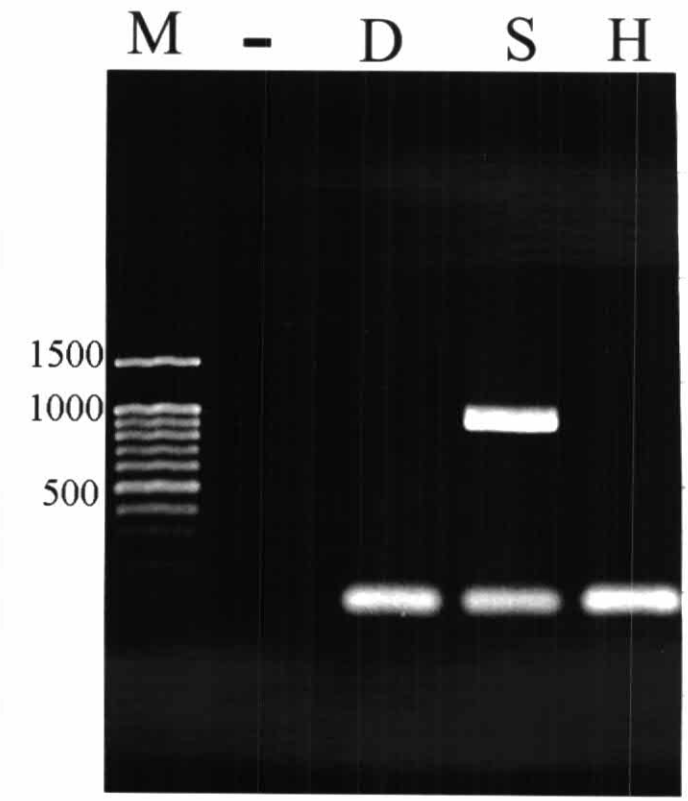


Figure 4.55 Secondary 5' GenomeWalk PCR product of *chg*-L separated in 1.5 % agarose gel. Lane M, -, D, S, and H is 100 bp DNA marker, negative control, *chg*-L/*Dra*I, *Ssp*I, and *Hae*III GenomeWalk product, respectively.

Because *chg*-L/*Dra*I primary PCR product was faint not enough for elution and *chg*-L/*Ssp*I may be same sequence as known sequence (*chg*-L D12-S4 see detail in appendix B and C), so primary *chg*-L/*Dra*I and *Hae*III were dilute 1:50 and used as template in secondary PCR amplification. DNA fragments at size 1.76 kb, 700 bp, and 300 bp were amplified from *chg*-L/*Dra*I primary product as shown in lane D of Fig. 4.56 B. DNA fragment at size 500 bp was amplified from *chg*-L/*Hae*III primary product as shown in lane H of Fig. 4.56 B.

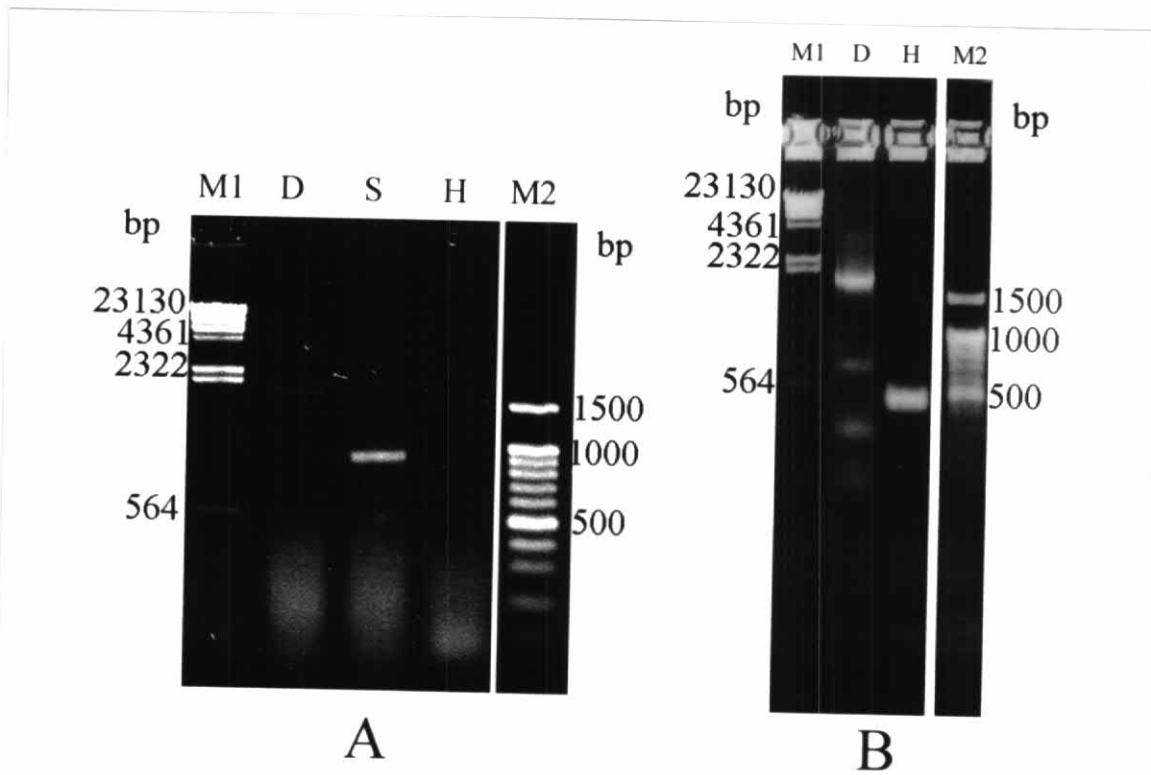


Figure 4.56 Primary (A) and secondary 5' GenomeWalk PCR product of *chg*-L (B) separated in 1.2 % agarose gel. Lane M1, M2, D, S, and H is λ*Hin*dIII, 100 bp DNA marker, *chg*-L/*Dra*I, *Ssp*I, and *Hae*III GenomeWalk product, respectively.

A DNA fragment at size 1.76 kb was sequenced (see in appendix B) and BLAST analysis (see in appendix C) found it homologous with chorion protein of *Sparus aurata* with E-value as 3e-34. Nucleotide sequence of *chg*-L D12/D20 was shown in Fig. 4.57. The sequence contained 5' upstream region at size 1,424 bp which contained 3 half-site EREs (GGTCA at nucleotide position -326 and -956, and TGACC at nucleotide position -1189). The full site sequence of ERE was not found.

CGACGGCCCGGGCAGGTAAATTAGATCCAGGGTGTCAGAACTGAGATCTGAAAATCCCCATTTGAGC
TCAGAGCAGAATAAACACAGTCGTTGTTGGTGCCCTGATGAAACAAAGGTTCAAAGGTTTTACTTGC
TTTCACAACCTGGATATCTTTTGGCTATTTGGTTATTTTTACAGTAATGAT*TGTTCT*TCATATATAT
GTATATATACAGTGATCAGGCATAACATTATGACCACTTGCCTAATATTGTGTAAGTCTCCATTGTG
CCTCAGTTGTGGCTCATCAGAGAGTGAACATGGGCCCTCTGAGGGTGCCCTGTGGTGTCTGTCAATA
TAATGTTGTTGGGGGGGGAGGGGGCTTTGGCTCCTAAGGGCTGAAAGGAGGGGCCTCTGTGAATCAG
GCTTGTTCCAGTCCATCCCAGAGACACTTGATCAGATTGGGAAGTTGTGAATTTGGAGGCCAGGTCA
ACACCTAGTACTGTTTGTCATGTTTTTTTTTAGTTGTTCCTAAACTATTTTTGTGTGTTTGTCTGTG
TCAGGCTGCATCCTGCTGGGGATGGCTGCTGCTTGCTATGGGGTGATGGTTGGGTTGTGGGGGGTAT
CTGGTCTATATATCAACCTTGCTTTGTGCACGTGAAGGATGGTTATGTAGAAAACATACCTACAATG
TGTTACTGTAGAGTCTACTAATGAAATACTGCAAATATTAAATAAATTAATACAATCTATTATGAAG
ATTTATGTAATATAGGCTTAAACAAATCAAATAAAGCTACGATAGAAAACAAACATTTAATGGAATA
TCAGAAGAGGAGGAGAGGAAGGATCAGCCCAAATTAAAAGTTTTGTCATGAGGTGTATGTGGCAGTA
ATATTATCACAATAATAAAACAAGATCTTCAGTCCAGTGATTTTTCCTCAGGTCTGAGCTGGTCCTA
AAGAATCGTAAATAAATAATATGGTAAACTTACTGCAGGTGAAGCTGACAACCTGCCCACCGTGGAG
CCATATCGACTGGCTGACATTCACAAAATGTACTTTGGCAAAACACTGAGTTCAGTCTGTCATGTCC
AGGGCAGGTAAACCTAACCCAGGGTCAACATAACTTTTTCTGCCTCTTTTTATTCTGTCTTTCTATA
TATCTTCAAGCTATTTGTAAAGATTTTTTAATATAATTTGTTTTCAGAAAATGTGTCTGTATTTATC
TCATCCCATAAGAAATAAAATGTAAATTGAGCTTTCAGTCAACAGCTGCATTTTATTGGCTACAGCT
AGGGCTTACGAGTCTGCTGAGGGATGGCCAACTGAAGGGGCGGGGAAGTAATGAAAATAAACAGGTT
CACAGCGTTATAAAAGTAGAAGCAAGTTCCCACCACTCTTACAGCACCTTGAGAATCTCTCAGATCG
CTTGTCACTGTGGAGCCATGGTGATGAAGTGGACTGCTGCTTGCCTTGTGGCACTGGCTCTATTTGC
CAGCGTCTGTGATGCTCAGTGGGGAGAGTACACGCCTTCAAAATATCAGAAACCTGCACCTCCTGTG
AAGCAAGAGCCCAAACAAGTGCCTCAAGACACTCAACAGCATAAGCAGACATTTGAAACACCACTTC
AATGGACATACCCTGAACCTCCCCCGCCTGAACCTGCGCCTGAAATACCATTTGAACCGCTACGTCC
TCAACCTGTTGCATCTGTTGCTGTTGAGTGCAGAGAGAATGATGCTCATGTGGAAGTCAGGAGGGAT
ATGTTTGGG

Figure 4.57 Nucleotide sequence of *chg*-L D12/D20. Red alphabet is start codon, blue label is 5' upstream region, and yellow label is half-sites EREs.

4.7 Cloning and characterization of transcriptional intermediary factor 2 (*tif2*) genes.

Partial sequence of *tif2* was amplified by PCR using *tif2* degenerate primer designed from conserved region in basic helix loop helix region, receptor interaction domain, and glutamine rich region of TIF2 from fish and other vertebrate species. PCR amplification with egg, testis, intestine, and liver 1st strand cDNA template of

L. subviridis but overall PCR product was amplified is other genes or transcript, so specific primer for amplification of *tif2* RID was designed from *tif2* cDNA sequence of zebrafish (*Danio rerio*) which reported in GenBank. DNA fragment at size 700 bp was amplified from adult head and visceral organ 1st strand cDNA template of normal *Danio rerio* as shown in Fig. 4.58.

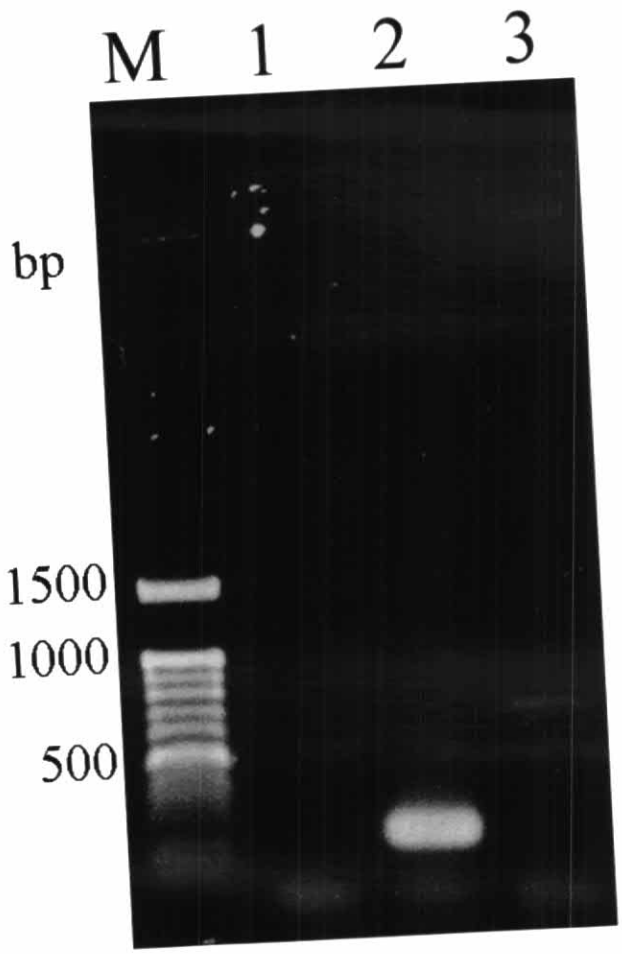


Figure 4.58 *Danio rerio tif2* RID PCR product separated in 1.2 % agarose gel. Lane M, 1, 2, and 3 is 100 bp DNA marker negative control, positive control, and *Danio rerio tif2* RID PCR product, respectively

4.8 Estrogen response of *chg*-L, *chg*-H, and *vtg*-3 genes in liver of *L. subviridis* detected by semi-quantitative RT-PCR

4.8.1 β-actin amplification

β-actin PCR product of *L. subviridis* at size 300 bp was amplified from liver (used β-actin primer of *Macrobrachium rosenbergii* (designed by Preechaphol) in cross-amplified) result shown in Fig. 4.59, after sequencing and BLAST analysis (see in appendix B and C) it homologous with β-actin of swamp eel (*Monopterus albus*) with E-value as 1e-117. Specific primer of *L. subviridis* β-actin was designed and used for internal control in semi-quantitative RT-PCR analysis.

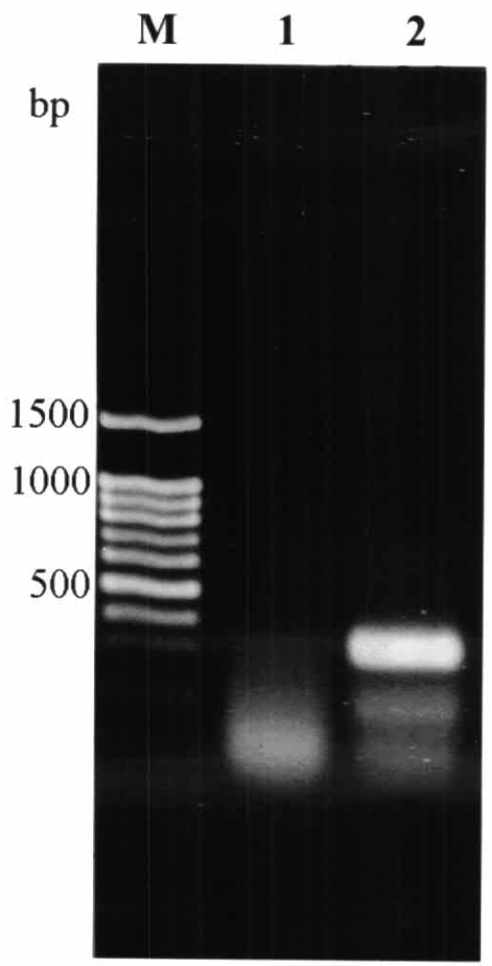


Figure 4.59 β-actin PCR product of *L. subviridis* separated in 1.2 % agarose gel. Lane M, 1, and 2 is 100 bp DNA marker, negative control, and β-actin PCR product, respectively.

β-actin PCR product were amplified from liver of all sample for used as control as shown in Fig. 4.60

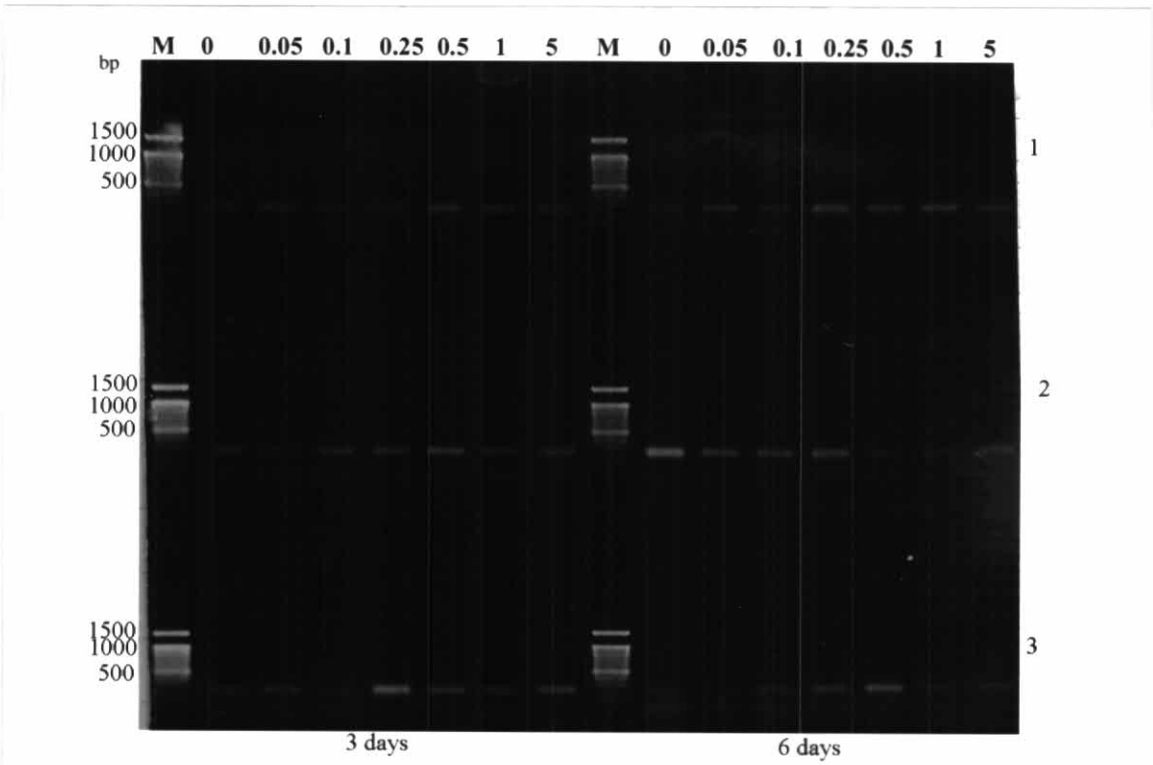


Figure 4.60 Liver β-actin PCR products of *L. subviridis* exposed with E_2 at dose 0-5 mg/kg for 3 and 6 days separated in 1.6 % agarose gel. Lane M is 100 bp DNA marker, No. of lane indicated dose of E_2 (mg/kg), left and right panel is 3 and 6 days exposed, respectively, up, middle, and low row is replicate 1, 2, and 3, respectively.

Expression level of β-actin gene in liver is constitutive in all sample used in experiment which indicated quality of template were acceptable (Fig. 4.60).

4.8.2 *chg*-L expression level in liver of *L. subviridis* after exposed with 17β-estradiol (estrogen).

chg-L PCR product were amplified from liver of *L. subviridis* exposed with E_2 at dose 0, 0.05, 0.1, 0.25, 0.5, 1, and 5 mg/kg for 3 and 6 days, results of amplification shown in Fig. 4.61.

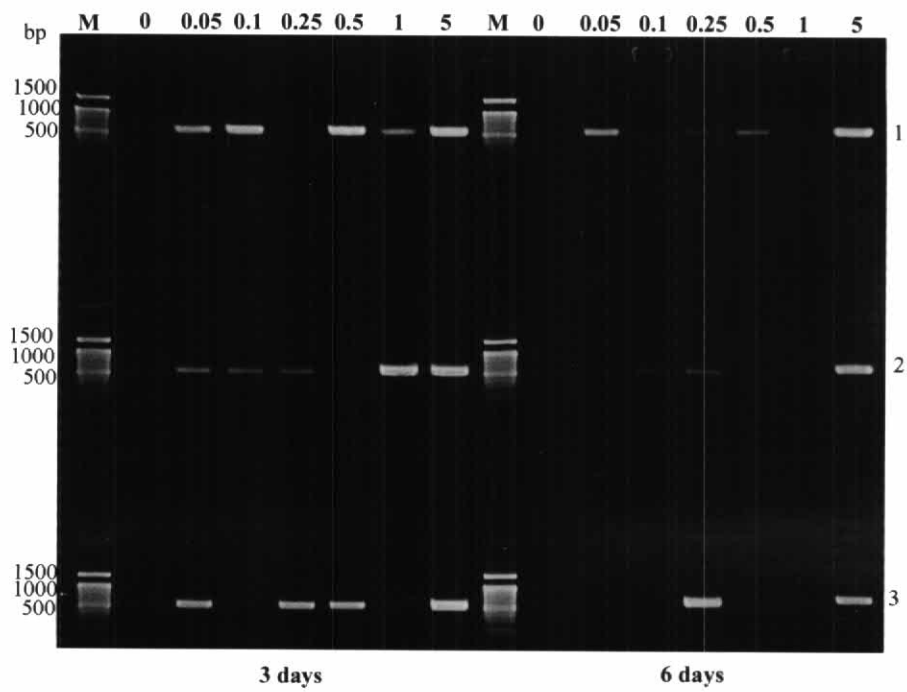
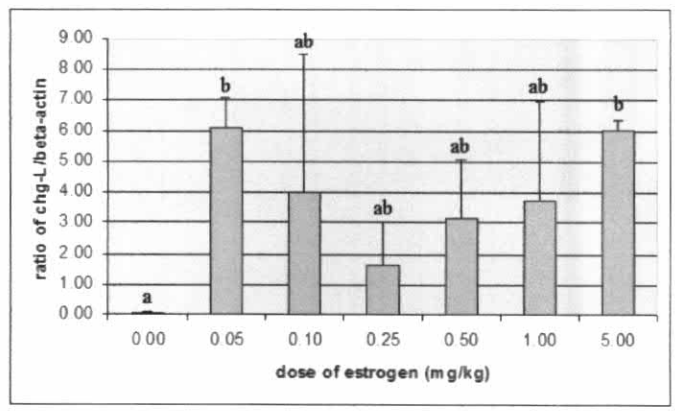
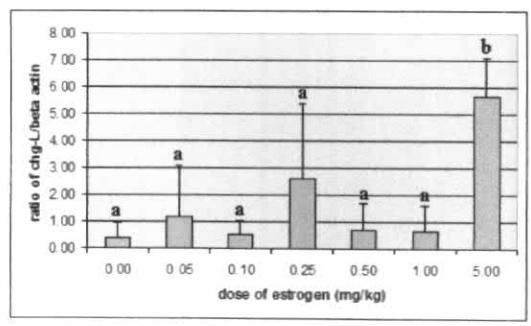


Figure 4.61 Liver *chg*-L PCR products of *L. subviridis* exposed with E_2 at dose 0-5 mg/kg for 3 and 6 days separated in 1.6 % agarose gel. Lane M is 100 bp DNA marker, No. of lane indicated dose of E_2 (mg/kg), left and right panel is 3 and 6 days exposed, respectively, row 1, 2, and 3 is replicate 1, 2, and 3, respectively.

Intensity of *chg*-L band were measured and normalized with actin (ratio of *chg*-L/*β* actin). mRNA expression level of *chg*-L in liver of *L. subviridis* increased significantly ($P < 0.05$) after exposed with estrogen at dose 0.05 and 5 mg/kg for 3 days but mRNA expression level of *chg*-L not changed when mullets exposed with E_2 at dose 0, 0.1, 0.25, 0.5, and 1 mg/kg (Fig. 4.62 A). mRNA expression level of *chg*-L in liver of *L. subviridis* increased significantly ($P < 0.05$) after exposed with estrogen at dose 5 mg/kg for 6 days but mRNA expression level of *chg*-L not changed when mullets exposed with E_2 at dose 0, 0.05, 0.1, 0.25, 0.5, and 1 mg/kg (Fig. 4.62 B).



A



B

Figure 4.62 mRNA expression level of *chg*-L gene in liver of *L. subviridis* which exposed estrogen by intraperitoneally injection at 3 days (A) at 6 days (B).

4.8.3 *chg*-H expression level in liver of *L. subviridis* after exposed with 17β-estradiol (estrogen).

chg-H PCR product were amplified from liver of *L. subviridis* exposed with E_2 at dose 0, 0.05, 0.1, 0.25, 0.5, 1, and 5 mg/kg for 3 and 6 days, results of amplification shown in Fig. 4.63.

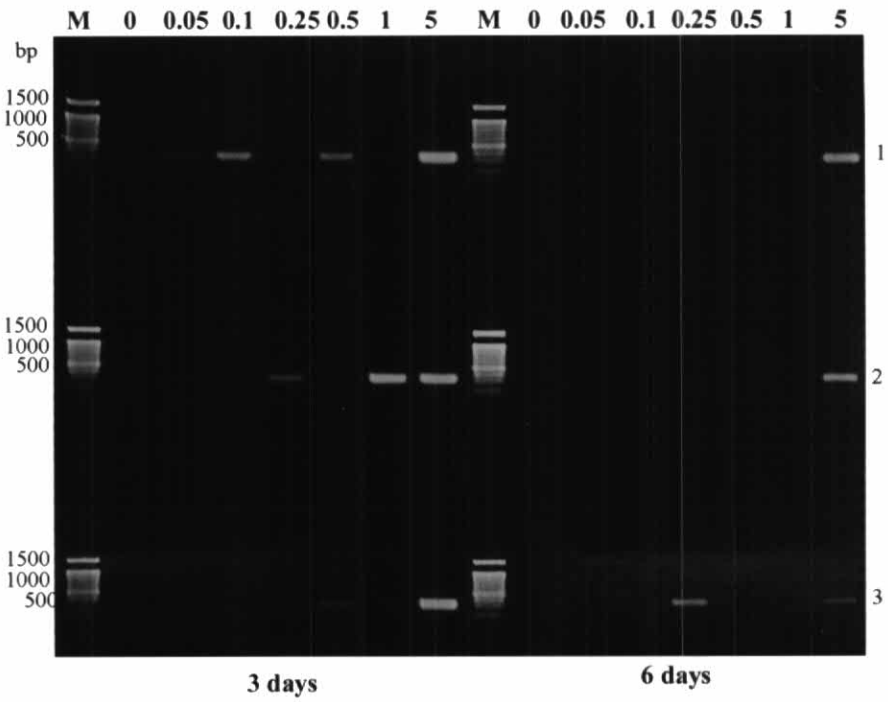
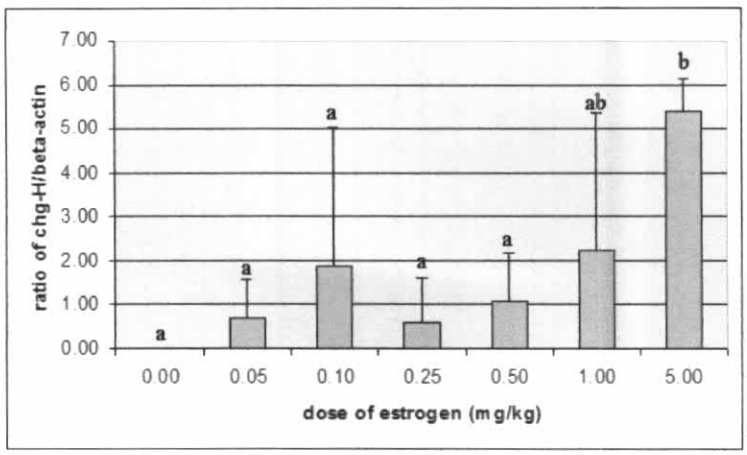
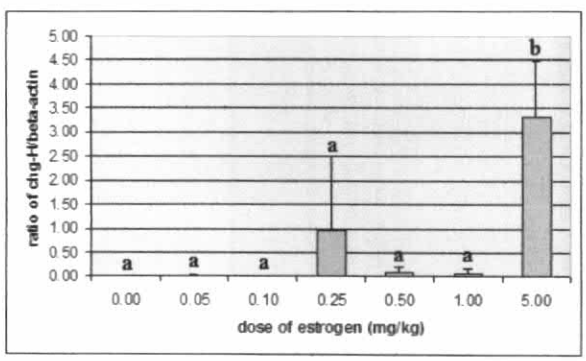


Figure 4.63 Liver *chg*-H PCR products of *L. subviridis* exposed with E_2 at dose 0-5 mg/kg for 3 and 6 days separated in 1.6 % agarose gel. Lane M is 100 bp DNA marker, No. of lane indicated dose of E_2 (mg/kg), left and right panel is 3 and 6 days exposed, respectively, row 1, 2, and 3 is replicate 1, 2, and 3, respectively.

Intensity of *chg*-H band were measured and normalized with actin (ratio of *chg*-H/*β* actin). mRNA expression level of *chg*-H in liver of *L. subviridis* increased significantly ($P < 0.05$) after exposed with estrogen at dose 5 mg/kg for 3 and 6 days but do not change when mullets exposed with E_2 at dose 0, 0.05, 0.1, 0.25, 0.5, and 1 mg/kg (Fig. 4.64)



A



B

Figure 4.64 mRNA expression level of *chg*-H gene in liver of *L. subviridis* which exposed estrogen by intraperitoneally injection at 3 days (A) at 6 days (B).

4.8.4 *vtg*-3 expression level in liver of *L. subviridis* after exposed with 17β-estradiol (estrogen).

vtg-3 PCR product were amplified from liver of *L. subviridis* exposed with E_2 at dose 0, 0.05, 0.1, 0.25, 0.5, 1, and 5 mg/kg for 3 and 6 days, results of amplification shown in Fig. 4.65 A and B, respectively.

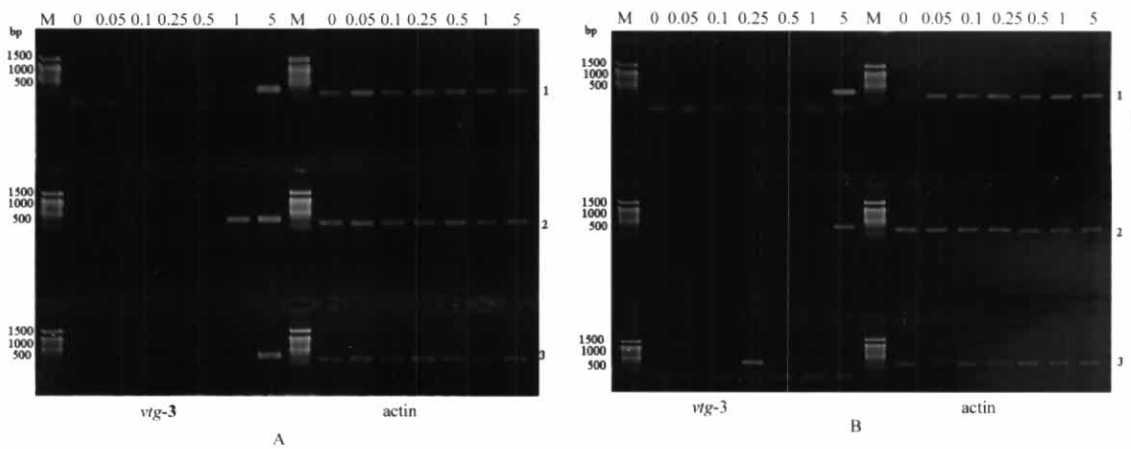
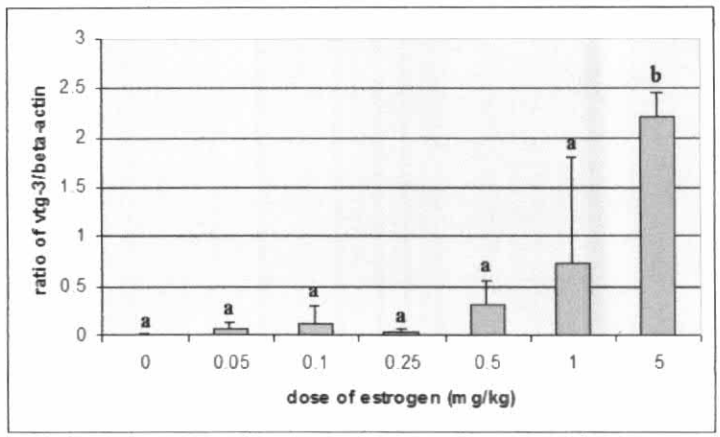


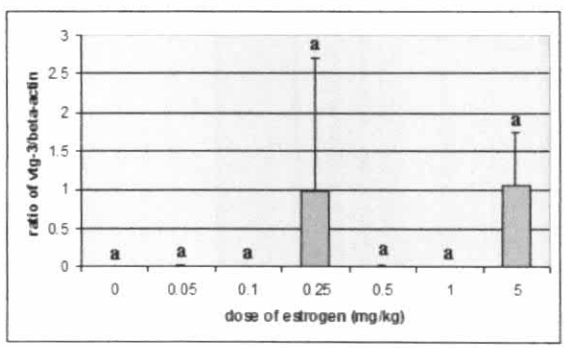
Figure 4.65 Liver *vtg*-3 and *β* actin PCR products of *L. subviridis* exposed with E_2 at dose 0-5 mg/kg for 3 days (A) and 6 days (B) separated in 1.6 % agarose gel. Lane M is 100 bp DNA marker, No. of lane indicated dose of E_2 (mg/kg), left and right panel is *vtg*-3 and *β* actin PCR products respectively, row 1, 2, and 3 is replicate 1, 2, and 3, respectively.

Intensity of *vtg*-3 band were measured and normalized with actin (ratio of *vtg*-3/*β* actin). mRNA expression level of *vtg*-3 in liver of *L. subviridis* increased significantly ($P < 0.05$) after exposed with estrogen at dose 5 mg/kg for 3 and 6 days but do not change when mullets exposed estrogen at dose 0, 0.05, 0.1, 0.25, 0.5, and 1 mg/kg. Results in Fig. 4.66 B shown mRNA expression level of *vtg*-3 in

liver of *L. subviridis* not changes after exposed estrogen at dose 0, 0.05, 0.1, 0.25, 0.5, 1, and 5 mg/kg for 6 days.



A



B

Figure 4.66 mRNA expression level of *vtg*-3 gene in liver of *L. subviridis* which exposed estrogen by intraperitoneally injection at 3 days (A) at 6 days (B).