2. Identification of polyhedrin gene of Thai BmNPV

2.1 Cloning and sequencing of Thai BmNPV *polh*

To study the sequence of Thai BmNPV *polh*, specific primers for BmNPV *polh* were designed. When specific primers were used, PCR products of the expected size of 1,448 bp were observed on gel (Figure 10). Purified PCR product was cloned into pGEM®-T vector. Recombinant clones were identified by white colony selection. The selected clones were then screened for clone harboring the BmNPV *polh*.

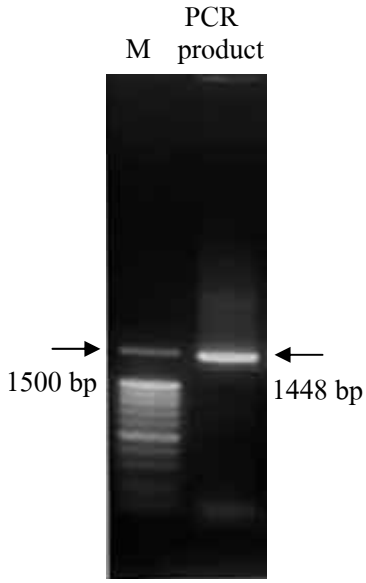


Figure 10 The expected PCR products of 1,448 bp amplified from the extracted DNA of Thai BmNPV.
M=DNA marker (100 bp+1.5kbp)

Two methods were used to screen clone harboring the BmNPV *polh* gene. The expected size of pGEM®-T vector with insert gene was shown in Figure 11 (a). Examples of clone selection for insert gene of BmNPV *polh* were illustrated in Figure 11 (b) and (c). At first, PCR was used to identify the recombinant clones. Figure 11 (b) showed three clones that were clarified by PCR and result indicated that only clone 1 revealed band of BmNPV *polh* of approximately 1,500 bp while clone 2 and clone 3 had no expected PCR products. Therefore, clone 1 may has an insert of *polh* gene. This result was rechecked by digestion recombinant clone with *Hin*dIII-*Sal*I. Figure 11 (c) showed *Hin*dIII-*Sal*I restriction pattern of the selected clone in which 3 bands with the size of 4.4 kb, 3 kb and 1.3 kb were observed. Fragment of 4.4 kb was the vector with insert gene, fragment of 3 kb was pGEM®-T vector and fragment of 1.3 kb was the insert gene.

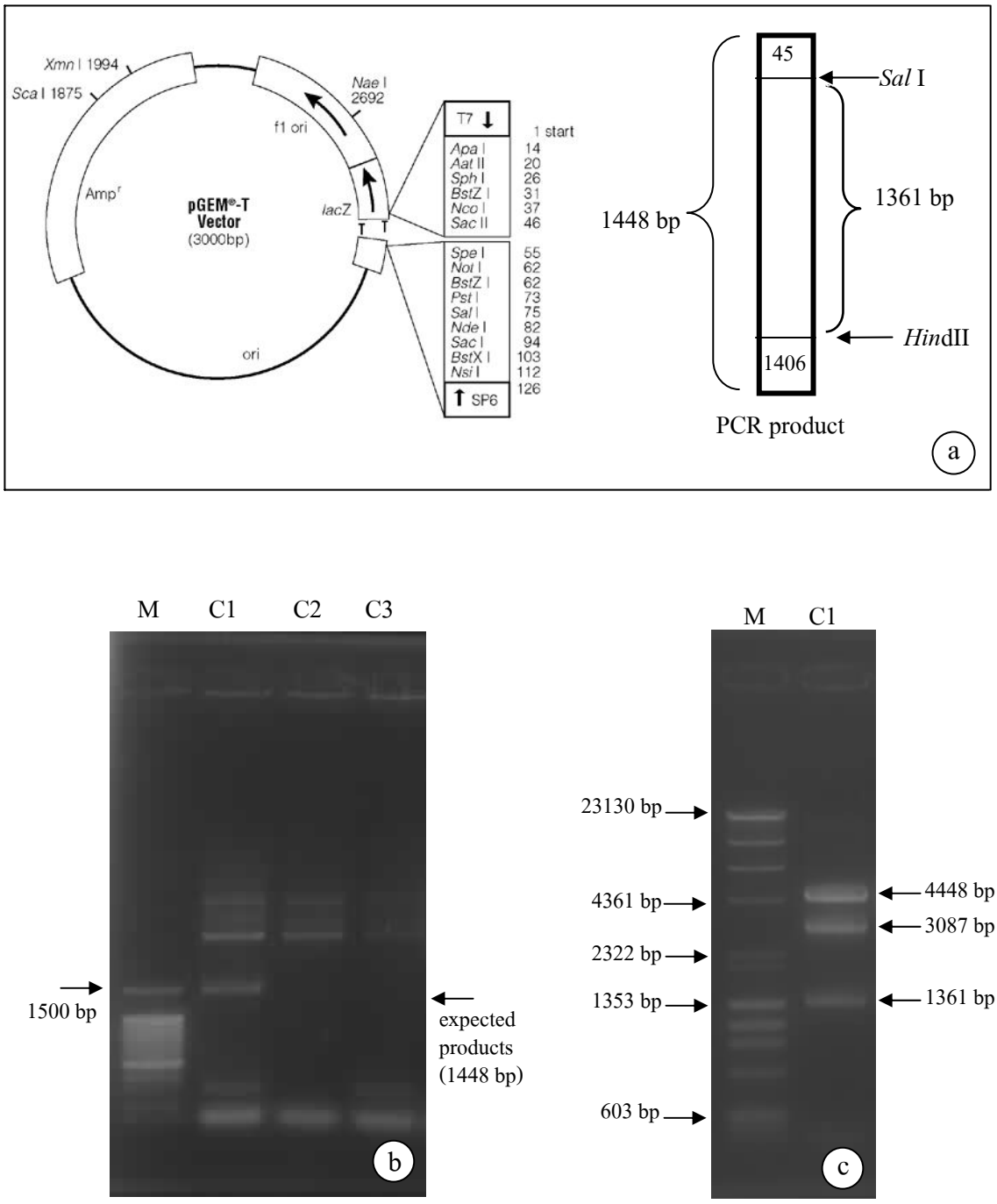


Figure 11 Screening of recombinant clones harboring the BmNPV *polh* gene.

a) Diagram of pGEM® -T vector and insert PCR product

b) Clone selection confirmed by PCR method, C=clones , M=marker 100 bp+1.5 Kb

c) Clone selection confirmed by *Hin*dIII-*Sal*I digestion, C=clones

M= marker :λ DNA- *Hin*dIII and ϕ X174 DNA-*Hae*III Mix

2.2 Nucleotide sequence of Thai BmNPV *polh*

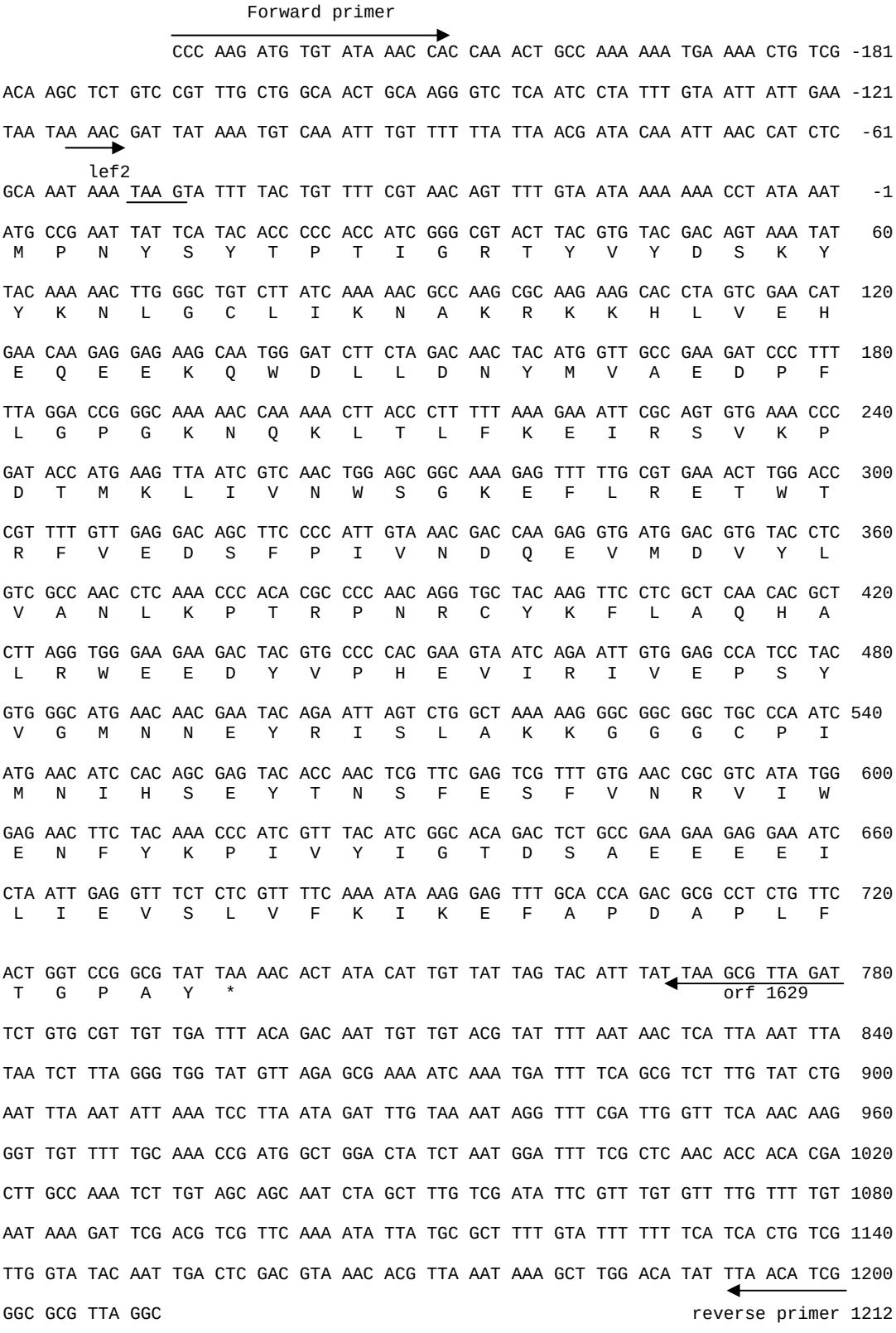
The PCR amplification products cloned into pGEM®-T vector showed the nucleotide sequence of 1,440-bp fragment of the Thai BmNPV DNA that contained the full-length coding region of the *polh*. The nucleotide sequence of Thai BmNPV *polh* and its flanking region was submitted to GenBank at electronic access (http://www.ncbi.nlm.nih.gov/) (Benson *et al*., 2000) and the accession number AY779044 was given as shown in Figure 12. The sequences and predicted amino acid were designed as shown in Figure 13. The *polh* ORF consisted of 735 nucleotides (not include stop codon) that encoded a polypeptide of 245 amino acids with the predicted molecular mass of 28.8 kDa. The submitted sequence contained 228 bp upstream of the translation initiation codon (ATG) and 474 bp downstream of the translation stop codon (TAA). Several characteristics of Thai BmNPV *polh* gene sequence were investigated. The immediate upstream sequence of the translation initiation site was AT rich and contained the unique conserved transcription start site TAAG motif which is similar to other baculovirus late gene promoters. There was no additional TAAG sequence in the sequence upstream of the initiation codon. As has been observed in all other hyperexpressed baculovirus late genes, there is a conserved sequence of 14 nucleotides upstream of the coding region with the consensus sequence of TAAATAAGTATTTT at position -42 to -56 (Leisy *et al*., 1986). DNA sequences similar to the consensus TATA and CAAT which represented important elements of eukaryotic gene promoters were observed at position -107 and -143. The canonical poly (A) signal AATAAA was present in the 3'end of the Thai BmNPV *polh* gene at position 1081. The Thai BmNPV *polh* ORF had the translation initiation codon, ATG, and the termination codon, TAA, as found in other baculoviruses (O'Reilly *et al*., 1992). The complete whole sequence of Thai BmNPV *polh* overlapped with *lef*-2 (late gene expression factor-2) in the 5' flanking region and *orf*1629 in the 3' flanking region. The *lef*-2 was located in the *polh* upstream region adjacent to the transcription start site in the same direction and the *orf*1629 was located in the *polh* downstream region in the reverse direction.

```
LOCUS       AY779044               1440 bp    DNA     linear   VRL 07-NOV-2004
DEFINITION  Bombyx mori nucleopolyhedrovirus from Thailand polyhedrin gene,
            complete cds.
ACCESSION   AY779044
VERSION     AY779044.1  GI:55247500
KEYWORDS    .
SOURCE      Bombyx mori NPV
  ORGANISM  Bombyx mori NPV
            Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
            Nucleopolyhedrovirus.
REFERENCE   1  (bases 1 to 1440)
  AUTHORS   Kaewwises,M., Attathom,T., Chaeychomsri,S. and Chowpongpang,S.
  TITLE     The Polyhedrin Gene of Thai Bombyx mori nucleopolyhedrovirus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1440)
  AUTHORS   Kaewwises,M., Attathom,T., Chaeychomsri,S. and Chowpongpang,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-2004) Biotechnology Research and Development
            Office, Pathum Thani 12110, Thailand
FEATURES             Location/Qualifiers
     source          1..1440
                     /organism="Bombyx mori NPV"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:271108"
                     /country="Thailand"
     CDS             229..966
                     /codon_start=1
                     /product="polyhedrin"
                     /protein_id="AAV48593.1"
                     /db_xref="GI:55247501"
                     /translation="MPNYSYTPTIGRTYVYDNKYYKNLGCLIKNAKRKKHLVEHEQEE
                     KQWDLLDNYMVAEDPFLGPGKNQKLTLFKEIRSVKPDTMKLIVNWSGKEFLRETWTRF
                     VEDSFPIVNDQEVMDVYLVANLKPTRPNRCYKFLAQHALRWEEDYVPHEVIRIVEPSY
                     VGMNNEYRISLAKKGGGCPIMNIHSEYTNSFESFVNRVIWENFYKPIVYIGTDSAEEE
                     EILIEVSLVFKIKEFAPDAPLFTGPAY"
ORIGIN
        1 cccaagatgt gtataaacca ccaaactgcc aaaaaatgaa aactgtcgac aagctctgtc
       61 cgtttgctgg caactgcaag ggtctcaatc ctatttgtaa ttattgaata ataaaacgat
      121 tataaatgtc aaatttgttt tttattaacg atacaaatta accatctcgc aaataaataa
      181 gtattttact gttttcgtaa cagttttgta ataaaaaaac ctataaatat gccgaattat
      241 tcatacaccc ccaccatcgg gcgtacttac gtgtacgaca ataaatatta caaaaacttg
      301 ggctgtctta tcaaaaacgc caagcgcaag aagcacctag tcgaacatga acaagaggag
      361 aagcaatggg atcttctaga caactacatg gttgccgaag atccctttt aggaccgggc
      421 aaaaaccaaa aacttaccct ttttaaagaa attcgcagtg tgaaacccga taccatgaag
      481 ttaatcgtca actggagcgg caaagagttt ttgcgtgaaa cttggacccg ttttgttgag
      541 gacagcttcc ccattgtaaa cgaccaagag gtgatggacg tgtacctcgt cgccaacctc
      601 aaacccacac gccccaacag gtgctacaag ttcctcgctc aacacgctct taggtgggaa
      661 gaagactacg tgccccacga agtaatcaga attgtggagc catcctacgt gggcatgaac
      721 aacgaataca gaattagtct ggctaaaaag ggcggcggct gcccaatcat gaacatccac
      781 agcgagtaca ccaactcgtt cgagtcgttt gtgaaccgcg tcatatggga gaacttctac
      841 aaacccatcg tttacatcgg cacagactct gccgaagaag aggaaatcct aattgaggtt
      901 tctctcgttt tcaaaataaa ggagtttgca ccagacgcgc ctctgttcac tggtccggcg
      961 tattaaaaca ctatacattg ttattagtac atttattaag cgttagattc tgtgcgttgt
     1021 tgatttacag acaattgttg tacgtatttt aataactcat taaatttata atctttaggg
     1081 tggtatgtta gagcgaaaat caaatgattt tcagcgtctt tgtatctgaa tttaaatatt
     1141 aaatccttaa tagatttgta aaataggttt cgattggttt caaacaaggg ttgtttttgc
     1201 aaaccgatgg ctggactatc taatggattt tcgctcaaca ccacacgact tgccaaatct
     1261 tgtagcagca atctagcttt gtcgatattc gtttgtgttt tgttttgtaa taaagattcg
     1321 acgtcgttca aaatattatg cgcttttgta ttttttttcat cactgtcgtt ggtatacaat
     1381 tgactcgacg taaacacgtt aaataaagct tggacatatt taacatcggg cgcgttaggc
```

Figure 12 Information of GenBank accession number AY779044, the submitted sequences of Thai BmNPV polyhedrin gene.

Source: http://www.ncbi.nih.govlentrez/viewer.fcgi?db=nucleotide&val=55247500

Figure 13 Nucleotide sequence of Thai BmNPV polyhedrin gene and its flanking regions. A total of 1,440 base pairs was submitted to GenBank (AY779044). The first nucleotide of the polyhedrin gene translational start signal ATG is given number 1. The predicted amino acid sequence is indicated by one-letter code and displayed below the nucleotide sequence. The sequence at nt -51 initiates the 5' end of the mRNA. The putative transcription initiation motif (TAAG) is underlined. The potential polyadenylation signal (ATTAAA) is indicated by overlining. The primer used in the primer extension assay is also shown and the arrow indicates the direction of extension.




```
                    CCC AAG ATG TGT ATA AAC CAC CAA ACT GCC AAA AAA TGA AAA CTG TCG -181

ACA AGC TCT GTC CGT TTG CTG GCA ACT GCA AGG GTC TCA ATC CTA TTT GTA ATT ATT GAA -121

TAA TAA AAC GAT TAT AAA TGT CAA ATT TGT TTT TTA TTA ACG ATA CAA ATT AAC CAT CTC  -61

GCA AAT AAA TAA GTA TTT TAC TGT TTT CGT AAC AGT TTT GTA ATA AAA AAA CCT ATA AAT   -1

ATG CCG AAT TAT TCA TAC ACC CCC ACC ATC GGG CGT ACT TAC GTG TAC GAC AGT AAA TAT   60
 M   P   N   Y   S   Y   T   P   T   I   G   R   T   Y   V   Y   D   S   K   Y

TAC AAA AAC TTG GGC TGT CTT ATC AAA AAC GCC AAG CGC AAG AAG CAC CTA GTC GAA CAT  120
 Y   K   N   L   G   C   L   I   K   N   A   K   R   K   K   H   L   V   E   H

GAA CAA GAG GAG AAG CAA TGG GAT CTT CTA GAC AAC TAC ATG GTT GCC GAA GAT CCC TTT  180
 E   Q   E   E   K   Q   W   D   L   L   D   N   Y   M   V   A   E   D   P   F

TTA GGA CCG GGC AAA AAC CAA AAA CTT ACC CTT TTT AAA GAA ATT CGC AGT GTG AAA CCC  240
 L   G   P   G   K   N   Q   K   L   T   L   F   K   E   I   R   S   V   K   P

GAT ACC ATG AAG TTA ATC GTC AAC TGG AGC GGC AAA GAG TTT TTG CGT GAA ACT TGG ACC  300
 D   T   M   K   L   I   V   N   W   S   G   K   E   F   L   R   E   T   W   T

CGT TTT GTT GAG GAC AGC TTC CCC ATT GTA AAC GAC CAA GAG GTG ATG GAC GTG TAC CTC  360
 R   F   V   E   D   S   F   P   I   V   N   D   Q   E   V   M   D   V   Y   L

GTC GCC AAC CTC AAA CCC ACA CGC CCC AAC AGG TGC TAC AAG TTC CTC GCT CAA CAC GCT  420
 V   A   N   L   K   P   T   R   P   N   R   C   Y   K   F   L   A   Q   H   A

CTT AGG TGG GAA GAA GAC TAC GTG CCC CAC GAA GTA ATC AGA ATT GTG GAG CCA TCC TAC  480
 L   R   W   E   E   D   Y   V   P   H   E   V   I   R   I   V   E   P   S   Y

GTG GGC ATG AAC AAC GAA TAC AGA ATT AGT CTG GCT AAA AAG GGC GGC GGC TGC CCA ATC 540
 V   G   M   N   N   E   Y   R   I   S   L   A   K   K   G   G   G   C   P   I

ATG AAC ATC CAC AGC GAG TAC ACC AAC TCG TTC GAG TCG TTT GTG AAC CGC GTC ATA TGG  600
 M   N   I   H   S   E   Y   T   N   S   F   E   S   F   V   N   R   V   I   W

GAG AAC TTC TAC AAA CCC ATC GTT TAC ATC GGC ACA GAC TCT GCC GAA GAA GAG GAA ATC  660
 E   N   F   Y   K   P   I   V   Y   I   G   T   D   S   A   E   E   E   E   I

CTA ATT GAG GTT TCT CTC GTT TTC AAA ATA AAG GAG TTT GCA CCA GAC GCG CCT CTG TTC  720
 L   I   E   V   S   L   V   F   K   I   K   E   F   A   P   D   A   P   L   F


ACT GGT CCG GCG TAT TAA AAC ACT ATA CAT TGT TAT TAG TAC ATT TAT TAA GCG TTA GAT  780
 T   G   P   A   Y   *                                           orf 1629

TCT GTG CGT TGT TGA TTT ACA GAC AAT TGT TGT ACG TAT TTT AAT AAC TCA TTA AAT TTA  840

TAA TCT TTA GGG TGG TAT GTT AGA GCG AAA ATC AAA TGA TTT TCA GCG TCT TTG TAT CTG  900

AAT TTA AAT ATT AAA TCC TTA ATA GAT TTG TAA AAT AGG TTT CGA TTG GTT TCA AAC AAG  960

GGT TGT TTT TGC AAA CCG ATG GCT GGA CTA TCT AAT GGA TTT TCG CTC AAC ACC ACA CGA 1020

CTT GCC AAA TCT TGT AGC AGC AAT CTA GCT TTG TCG ATA TTC GTT TGT GTT TTG TTT TGT 1080

AAT AAA GAT TCG ACG TCG TTC AAA ATA TTA TGC GCT TTT GTA TTT TTT TCA TCA CTG TCG 1140

TTG GTA TAC AAT TGA CTC GAC GTA AAC ACG TTA AAT AAA GCT TGG ACA TAT TTA ACA TCG 1200

GGC GCG TTA GGC                                                 reverse primer 1212
```

The orientation of *polh*, *lef*-2 and *orf*1629 in BmNPV are the same as those found in AcNPV genome. The position of *orf1629* gene in HaNPV, SeNPV and SlNPV is next to *polh* similar to BmNPV but the position of *lef*-2 is different. In HaNPV, SeNPV and SlNPV, *lef*-2 is not located in *polh* upstream. In BmNPV *lef*-2 is essential for both viral DNA replication and late gene expression (Sriram and Gopinathan, 1998) and the *orf*1629 is essential for BmNPV viability (Je *et al*., 2001).

Flanking nucleotides of Thai BmNPV *polh* and BmNPV *polh* fragment which was used to design the primers, were compared. The major difference between the two isolates was the deletion of eight nucleotides in the upstream region of Thai BmNPV *polh* ORF while the rest of the nucleotide sequences were almost identical (Figure 14). Since the eight nucleotides are located upstream of the *polh* promoter, it might have no effect on the expression of *polh* promoter. A study on recombinant expression indicated that the suitable multiple cloning site (MCS) is located immediately downstream of the *polh* promoter (Acharya *et al*., 2002). In addition, there was a study on series of deletion in the upstream region of BmNPV *polh*s *(lef*-2, *orf*327, *orf*453 and *bro-e*) and the results revealed that the upstream region of *polh* has no effect on expression of *polh* promoter (Acharya and Gopinathan, 2001).

```
Thai                  CCCAAGATGTGTATAAACCACCAAACTGCCAAAAAATGAAAACTGTCGACAAGCTC -173
Japanese 1 (T3)       CCCAAGATGTGTATAAACCACCAAACTGCCAAAAAATGAAAACTGTCGACAAGCTC -181
                                                                  lef-2

Thai              TGTCCGTTTGCTGGCAACTGCAAGGGTCTCAATCCTATTTGTAATTATTGAATAATAAAA -103
Japanese 1(T3)    TGTCCGTTTGCTGGCAACTGCAAGGGCCTCAATCCTATTTGTAATTATTGAACAATAAAA -121

Thai              CGATTATAAATGTCAAATTTGTTTTTTATTAACGATACAAAT--------TAACCATCTC  -61
Japanese 1(T3)    CAATTATAAATGTCAAATTTGTTTTTTATTAACGATACAAATGGAAATAATAACCATCTC  -61

Thai              GCAAATAAATAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAAT   -1
Japanese 1 (T3)   GCAAATAAATAAGTATTTTACTGTTTTCGTAACAGTTTTGTAATAAAAAAACCTATAAAT   -1
```

Figure 14 Comparison of flanking nucleotide of Thai BmNPV *polh* (AY779044) and Japanase 1 (T3) (L33180). The different nucleotide sequences are shaded. The deletion region in Thai BmNPV is indicated by dashes.

2. 3 Comparison of Thai BmNPV *polh* with other BmNPV *polh*s

The nucleotides of Thai BmNPV *polh* were compared with those of five other BmNPV; Japanese 1 (T3) (Gomi *et al*., 1999), Japanese 2 (Maeda *et al*., 1985), Korean (K1), Chinese and Canadian (Iatrou *et al*., 1985). All BmNPV *polh* ORFs contained 735 nucleotides that encoded a polypeptide of 245 amino acids. The coding portion of the *polh* of all BmNPVs was not interrupted by intervening sequences. The nucleotide and amino acid sequences of the BmNPV *polh* of Thai and Japanese 1 (T3) isolate were identical. Percentage identity of the BmNPV *polh* sequence of the Thai isolate compared with the isolates from Japan (Japanese 2), Korea, China, and Canada were 99.9%, 99.7%, 99.6% and 98.8%, respectively, and percent identity of Thai BmNPV Polh amino acid sequence compared with those of the above-mentioned BmNPV isolate were 100%, 99.6%, 99.6%, 99.6% and 96.7%, respectively (Figure 15).

Percentage nucleotide sequence identity								
	1	2	3	4	5	6	varieties	
1		100	99.9	99.7	99.6	98.4	1	Thai
2	100		99.9	99.7	99.6	98.4	2	Japanese 1(T3)
3	99.6	99.6		99.6	99.5	98.5	3	Japanese 2
4	99.6	99.6	99.2		99.3	98.1	4	Korean(K1)
5	99.6	99.6	99.2	99.2		98.0	5	Chinese
6	96.7	96.7	97.1	96.3	96.3		6	Canadian
	1	2	3	4	5	6		
Percentage amino acid sequence identity								

Figure 15 Nucleotide and amino acid sequence identity of BmNPV *polh*s. Percentage nucleotide identity is denoted above the diagonal and percentage amino acid identity is below. Sequence data of BmNPV *polh*s were retrieved from GenBank: 1) Thai (AY779044); 2) Japanese 1 (T3) (L33180); 3) Japanese 2 (M30925); 4) Korean (K1) (U75359); 5) Chinese (X63614) and 6) Canadian (M100430).

Five amino acid sequences of BmNPV Polhs were aligned with Thai BmNPV Polh and results showed variation of the amino acid sequence in the N-terminus (Figure 16). In comparison between the Polh amino acid sequence of Thai BmNPV with those of the Japanese 2, Korean (K1) and Chinese, there was only one amino acid difference. An amino acid substitution was observed at M^{56}/V in Japanese 2, at P^{5}/S in Korean, and at Q^{57}/E in Chinese Polh. Among the five isolates of *Bm*NPV, the Canadian isolate showed the most difference from the Thai *Bm*NPV Polh. There were 8 amino acid differences: substitution at position N^{7}/T, G^{26}/C, I^{38}/V, K^{42}/Q, V^{75}/I , N^{77}/S, D^{147}/E and M^{156}/V. The amino acid sequence beyond position 156 of BmNPV Polh was conserved in all BmNPVs studied. Comparison of the nucleotide and amino acid sequences among BmNPVs indicated that BmNPVs found in the Asian countries (Thailand, Japan, Korea and China) were more closely related than the isolate from the North American country (Canada).

```
           Thai MPNYSYTPTIGRTYVYDNKYYKNLGCLIKNAKRKKHLVEHEQEEKQWDLLDNYMVAEDPFLGPGK  65
Japanese 1 (T3) MPNYSYTPTIGRTYVYDNKYYKNLGCLIKNAKRKKHLVEHEQEEKQWDLLDNYMVAEDPFLGPGK  65
    Japanese 2  MPNYSYTPTIGRTYVYDNKYYKNLGCLIKNAKRKKHLVEHEQEEKQWDLLDNYMVAEDPFLGPGK  65
    Korean (K1) MPNYPYTPTIGRTYVYDNKYYKNLGCLIKNAKRKKHLVEHEQEEKQWDLLDNYMVAEDPFLGPGK  65
       Chinese  MPNYSYTPTIGRTYVYDNKYYKNLGCLIKNAKRKKHLVEHEQEEKQWDLLDNYMVAQDPFLGPGK  65
      Canadian  MPNYSYNPTIGRTYVYDNKYYKNLGGLIKNAKRKKHLIEHEKEEKQWDLLDNYMVAEDPFLGPGK  65
     Consensus  MPNYSYTPTIGRTYVYDNKYYKNLGCLIKNAKRKKHLVEHEQEEKQWDLLDNYMVAEDPFLGPGK  65

           Thai NQKLTLFKEIRSVKPDTMKLIVNWSGKEFLRETWTRFVEDSFPIVNDQEVMDVYLVANLKPTRPN 130
Japanese 1 (T3) NQKLTLFKEIRSVKPDTMKLIVNWSGKEFLRETWTRFVEDSFPIVNDQEVMDVYLVANLKPTRPN 130
     Japanese 2 NQKLTLFKEIRSVKPDTMKLIVNWSGKEFLRETWTRFVEDSFPIVNDQEVMDVYLVANLKPTRPN 130
    Korean (K1) NQKLTLFKEIRSVKPDTMKLIVNWSGKEFLRETWTRFVEDSFPIVNDQEVMDVYLVANLKPTRPN 130
        Chinese NQKLTLFKEIRSVKPDTMKLIVNWSGKEFLRETWTRFVEDSFPIVNDQEVMDVYLVANLKPTRPN 130
      Canadian NQKLTLFKEVRNVKPDTMKLIVNWSGKEFLRETWTRFVEDSFPIVNDQEVMDVYLVANLKPTRPN 130
     Consensus NQKLTLFKEIRSVKPDTMKLIVNWSGKEFLRETWTRFVEDSFPIVNDQEVMDVYLVANLKPTRPN 130

           Thai RCYKFLAQHALRWEEDYVPHEVIRIVEPSYVGMNNEYRISLAKKGGGCPIMNIHSEYTNSFESFV 195
Japanese 1 (T3) RCYKFLAQHALRWEEDYVPHEVIRIVEPSYVGMNNEYRISLAKKGGGCPIMNIHSEYTNSFESFV 195
     Japanese 2 RCYKFLAQHALRWEEDYVPHEVIRIMEPSYVGMNNEYRISLAKKGGGCPIMNIHSEYTNSFESFV 195
    Korean (K1) RCYKFLAQHALRWEEDYVPHEVIRIVEPSYVGMNNEYRISLAKKGGGCPIMNIHSEYTNSFESFV 195
        Chinese RCYKFLAQHALRWEEDYVPHEVIRIVEPSYVGMNNEYRISLAKKGGGCPIMNIHSEYTNSFESFV 195
      Canadian RCYKFLAQHALRWDEDYVPHEVIRIMEPSYVGMNNEYRISLAKKGGGCPIMNIHSEYTNSFESFV 195
     Consensus RCYKFLAQHALRWEEDYVPHEVIRIVEPSYVGMNNEYRISLAKKGGGCPIMNIHSEYTNSFESFV 195

           Thai NRVIWENFYKPIVYIGTDSAEEEEILIEVSLVFKIKEFAPDAPLFTGPAY               245
Japanese 1 (T3) NRVIWENFYKPIVYIGTDSAEEEEILIEVSLVFKIKEFAPDAPLFTGPAY               245
     Japanese 2 NRVIWENFYKPIVYIGTDSAEEEEILIEVSLVFKIKEFAPDAPLFTGPAY               245
    Korean (K1) NRVIWENFYKPIVYIGTDSAEEEEILIEVSLVFKIKEFAPDAPLFTGPAY               245
       Chinese  NRVIWENFYKPIVYIGTDSAEEEEILIEVSLVFKIKEFAPDAPLFTGPAY               245
      Canadian  NRVIWENFYKPIVYIGTDSAEEEEILIEVSLVFKIKEFAPDAPLFTGPAY               245
     Consensus  NRVIWENFYKPIVYIGTDSAEEEEILIEVSLVFKIKEFAPDAPLFTGPAY               245
```

Figure 16 Alignment of the deduced amino acids of Thai BmNPV *polh* (AY779044) with other BmNPV *polhs*; Japanese 1 (T3) (L33180), Japanese 2 (M30925), Korean (K1)(U75359), Chinese (X63614) and Canadian (M100430). The alignments of five BmNPV *polh*s were made using the Clustal method, MegAlign program of DNASTAR. The different amino acids are shaded.

Previous study on the Canadian and Russian BmNPV Polh indicated that there were 8 amino acid differences (Iatrou *et al*., 1985). The homology of other very late gene, *p10*, of BmNPV from different isolates has been reported. Among six BmNPV isolates, four from Korea and two from China, two of them had the identical P10 amino acid sequences while the rest had 8-14 amino acid differences from each other (Hong *et al.*, 2000). The percentage identity of five BmNPV *p10* from India, China, Japan (2 isolates) and Taiwan was higher than 95.7% (Palhan and Gopinathan, 2000). Functional studies on *p10* and *polh* of BmNPV have suggested *polh* is more conserved than *p10* since *polh* showed less differences and high percent identity than *p10*.

A phylogenetic tree was generated from the deduced amino acid sequence of BmNPV *polh* using DNASTAR (MegAlign) (Figure 17). The results showed that BmNPV could be divided into two groups. The Asian group composed of BmNPV isolates from Thailand, Japan 1 (T3), Japan 2, Korea (K1) and China. The North American group composed of BmNPV isolate from Canada.

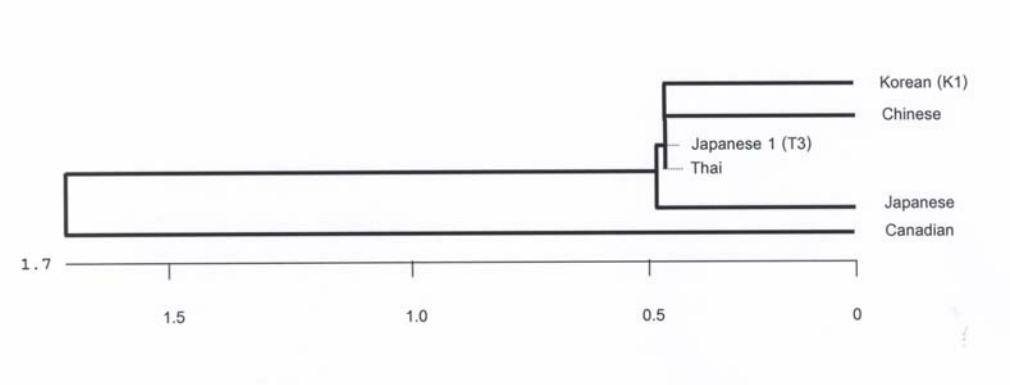


Figure 17 A phylogenetic tree of 6 BmNPV *polhs* (Thai: AY779044, Japanese 1 (T3): L33180, Japanese 2: M30925, Korean (K1): U75359, Chinese: X63614 and Canadian: M100430) constructed from amino acid sequences. The tree was generated by Clustal method, MegAlign program of DNASTAR.

Silkworm has been reared for centuries and has become a domestic animal species. Due to this life history, wings of silkworm have shortened and the silk moth can not longer fly. This immobility feature has restricted them to a confined region. Thus, it was believed that silkworms

and their NPVs may originally occur in the same place and during evolution they distributed and evolved individually. The limited vagility of the silkworm, biogeographic variation was observed between different isolates of NPVs in this species as demonstrated by phylogeographic differences between the Canadian and Asian isolates. However, the percentage identity of all BmNPV *polh* nucleotide sequences was greater than 98% which indicated that *polh* was a highly conserved gene.

2.4 Comparison of Thai BmNPV *polh* with other NPV *polh*s

Thai BmNPV *polh* nucleotides and amino acid sequences were compared with those of AcNPV (Ayres *et al*., 1994), HaNPV (Chen *et al*., 2001), SeNPV (IJkel *et al*., 1999) and SlNPV (Pang *et al*., 2001). AcNPV is a prototype of Baculovirus, and HaNPV, SeNPV and SlNPV are the viruses of economically important insect pests in Thailand. The percentage identities of Thai BmNPV *polh* gene sequence as compared with those of AcNPV, HaNPV, SeNPV and SlNPV were 75.2%, 65.3%, 70.1% and 65.9%, respectively, and percentage identities of the Polh sequence were 86.1%, 81.6%, 78.0%, and 79.6%, respectively (Figure 18).

Percentage nucleotide sequence identity							
	1	2	3	4	5		Virus
1		75.2	65.3	70.1	65.9	1	BmNPV
2	86.1		67.2	72.6	72.1	2	AcNPV
3	81.6	84.6		70.1	76.1	3	HaNPV
4	78.0	81.3	83.7		70.2	4	SeNPV
5	79.6	83.7	82.5	83.8		5	SlNPV
	1	2	3	4	5		
Percentage amino acid sequence identity							

Figure 1: Nucleotide and amino acid sequence identity of NPV *polhs*. Percentage nucleotide identity is denoted above the diagonal and percentage amino acid identity is below. Sequence data of NPV *polhs* were retrieved from GenBank: 1) BmNPV (AY779044); 2) AcNPV (NC_001623); 3) HaNPV (NC_002654); 4) SeNPV (NC_002169) and 5) SlNPV NC_003102).

ORF of BmNPV *polh* encoded a polyhedrin protein containing 245 amino acids similar to *polh* of AcNPV and HaNPV while the *polh* of SeNPV and SlNPV contained 246 and 249 amino acids, respectively. The amino acid sequence alignment of BmNPV Polh with other NPVs demonstrated that differences occurred more in the N-terminus than C-terminus (Figure 19). In HaNPV, the amino acid of Polh at position 40 is deleted. There are many substitution positions that make amino acids of BmNPV Polh differs from other NPV Polhs such as V^{30}/L, L^{50}/W, M^{128}/L, R^{129}/K, C^{147}/W, D^{148}/E, P^{149}/E, L^{186}/I, I^{199}/V and L^{226}/I. Amino acid contents of the putative BmNPV Polh indicated that it is rich in acidic amino acid residues, such as glutamic acid, especially at position 220-223 where four glutamic acids align consecutively. It is relative poor in cystein, tryptophan and histidine.

```
    BmNPV    ----MPNYSYTPTIGRTYVYDNKYYKNLGCLIKNAKRKKHLVEHEQEEKQWDLLDNYMVAEDPFL  65
    AcNPV    ----MPDYSYRPTIGRTYVYDNKYYKNLGAVIKNAKRKKHFAEHEIEEATLDPLDNYLVAEDPFL  65
    HaNPV    ---MYTRYSYSPTLGKTYVYDNKYFKNLGAVIKMPTQEH-LEEHEHEERNLDSLDKYLVAEDPFL  65
    SeNPV    ---MYTRYSYNPALGRTYVYDNKFYKNLGSVIKNAKRKEHLLQHEIEERTLDPLERYVVAEDPFL  65
    SlNPV    MYSRYSAYNYSPHLGKTYVYDNKYYKNLGHVIKNAKRKHDALEREADERELDHLDKYLVAEDPFM  65
Consensus        YT YSYSPTLGRTYVYDNKYYKNLGAVIKNAKRK HLLEHE EER LD LDKYLVAEDPFL  65

    BmNPV    GPGKNQKLTLFKEIRSVKPDTMKLIVNWSGKEFLRETWTRFVEDSFPIVNDQEVMDVYLVANLKP 130
    AcNPV    GPGKNQKLTLFKEIRNVKPDTMKLVVGWKGKEFYRETWTRFMEDSFPIVNDQEVMDVFLVVNMRP 130
    HaNPV    GPGKNQKLTLFKEIRSVKPDTMKLVVNWSGREFLRETWTRFMEDSFPIVNDQEIMDVFLSVNMRP 130
    SeNPV    GPGKNQKLTLFKEIRIVKPDTMKLVVNWSGKEFLRETWTRFMEDSFPIVNDQEIMDVFLVINMRP 130
    SlNPV    GPGKNQKLTLFKEIRNVKPDTMKLIVNWNGKEFLRETWTRFMEDSFPIVNDQEVMDVFLVVNMRP 130
Consensus    GPGKNQKLTLFKEIR VKPDTMKLVVNWSGKEFLRETWTRFMEDSFPIVNDQEVMDVFLVVNMRP 130

    BmNPV    TRPNRCYKFLAQHALRWEEDYVPHEVIRIVEPSYVGMNNEYRISLAKKGGGCPIMNIHSEYTNSF 195
    AcNPV    TRPNRCYKFLAQHALRCDPDYVPHDVIRIVEPSWVGSNNEYRISLAKKGGGCPIMNLHSEYTNSF 195
    HaNPV    TKPNRCYRFLAQHALRCDPDYIPHEVIRIVEPSYVGSNNEYRISLAKKYGGCPVMNLHAEYTNSF 195
    SeNPV    TRPNRCFRFLAQHALRCDPDYVPHEVIRIVEPVYVGTNNEYRISLAKKGGGCPVMNLHSEYTNSF 195
    SlNPV    TRPNRCFRFLAQHALRCDPEYVPHDVIRIVEPSYVGTNNEYRISLAKKGGGCPVMNLHAEYTTSF 195
Consensus    TRPNRCYRFLAQHALRCDPDYVPHEVIRIVEPSYVGSNNEYRISLAKKGGGCPVMNLHSEYTNSF 195

    BmNPV    ESFVNRVIWENFYKPIVYIGTDSAEEEEILIEVSLVFKIKEFAPDAPLFTGPAY            249
    AcNPV    EQFIDRVIWENFYKPIVYIGTDSAEEEEILLEVSLVFKVKEFAPDAPLFTGPAY            249
    HaNPV    EDFITNVIWENFYKPIVYVGTDSAEEEEILLEVSLIFKIKEFAPDAPLYTGPAY            249
    SeNPV    EEFINRVIWENFYKPIVYVGTDSGEEEEILLELSLVFKIKEFAPDAPLYNGPAY            249
    SlNPV    ESFIDKVIWYNFYKPIVYVGTDSAEEEEILLEVSLVFKIKEFAPDAPLYTGPAY            249
Consensus    E FI RVIWENFYKPIVYVGTDSAEEEEILLEVSLVFKIKEFAPDAPLYTGPAY            249
```

Figure 19 Alignment of the deduced amino acids of Thai *Bm*NPV *polh* (AY779044) with other NPV *polhs*; AcNPV (NC_001623), HaNPV (NC_002654), SeNPV(NC_002169) and SlNPV(NC_003102). The alignment of five NPV Polhs were made using the Clustal method, MegAlign program of DNASTAR. The different amino acids are shaded.

In general, percentage identity of NPV amino acid sequence was higher than nucleotide sequence which suggests that the genetic code is degenerated and the same amino acid is encoded by more than one codon. Previously, Chou *et al*. (1996) reported the percentage identities of amino acid sequence of the BmNPV Polh as compared to that of AcNPV, SeNPV, and SlNPV were 86%, 82% and 80%, respectively, which is similar to the results obtained from this study.

Phylogenetic tree of NPV Polhs showed that Thai BmNPV Polh was distinct from other NPVs. Based on multiple sequence alignment (MegAlign) analysis of the Polh, BmNPV separated from the group of AcNPV, HaNPV, SeNPV and SlNPV (Figure 20).

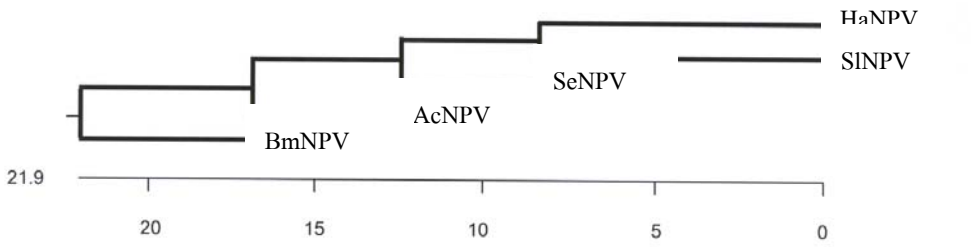


Figure 20 A phylogenetic tree of five NPV Polhs (BmNPV:AY779044, AcNPV:NC_001623, HaNPV: NC_002654, SeNPV:NC_002169 and SlNPV:NC_003102) constructed from amino acid sequences. The tree was generated by Clustal method, MegAlign program of DNASTAR.

Nucleopolyhedrovirus clades were first identified by Zanotto *et al.* (1993). Based on the Polh, they constructed a phylogenetic tree of baculoviruses in which NPVs were divided into Group I and Group II. Subsequently, Bulach *et al*. (1999) supported these clades and revealed other subclades under Group II by analyzing the Polh and DNA polymerase. Both Zanotto *et al.* (1993) and Bulach *et al*. (1999) grouped BmNPV and AcNPV into Group I and HaNPV, SeNPV and SlNPV into Group II.

Harrison and Bonning (2003) constructed a phylogenetic tree of Polh sequence of many lepidopteran NPVs including BmNPV, AcNPV, HaNPV, SeNPV and SlNPV. They grouped HaNPV, SeNPV and SlNPV into Group II and BmNPV into Group I of the proposed tree of Zanotto *et al*. (1993). The AcNPV polyhedrin was put on a branch outside of the clade containing the other members of Group I. This could suggest that AcNPV may acquire its *polh* gene by recombination with another virus that is not closely related to other NPVs in Group I. In addition, Jehel (2004) used the Hidden Markov Model to explain that AcNPV *polh* is a chimeric gene which consists of a mosaic of the genome of Group I and II NPVs. From these results AcNPV can be grouped in both Group I and Group II depending on the method used for the analysis. Since several reports revealed that adding the AcNPV *polh* resulted in distortion and instability to the *polh* gene tree, therefore, many other genes were recently employed for phylogenetic analysis of baculovirus (Herniou *et al*., 2004). However, for simple molecular analysis, *polh* is still useful because a great number of *polh* gene sequences are available in GenBank (Lange *et al*., 2004).