RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2011.37.205

ScienceAsia 37 (2011): **205–213**

Analysis of the immediate early-1 gene sequence of Thai *Bombyx mori* nucleopolyhedrovirus and its usefulness for identifying the subgroups of group II nucleopolyhedroviruses

Banthari Chotimanothum[a], **Thanaporn Pinpart**[b], **Lertluk Ngernsiri**[c,*]

[a] Bioscience Program, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand
[b] National Laboratory Animal Center, Mahidol University, Nakhon Pathom 73170, Thailand
[c] Department of Genetics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

*Corresponding author, e-mail: khawlln@yahoo.com

Received 3 Feb 2011
Accepted 11 Apr 2011

ABSTRACT: The immediate early 1 (*ie-1*) gene is essential for DNA replication of nucleopolyhedrosis virus. Here, the gene *ie-1* from nucleopolyhedrovirus (Nanglai-BmNPV) infecting a local silkworm (*Bombyx mori*), Nanglai strain, was cloned and sequenced. The gene contained a 1755 bp open reading frame encoding a deduced protein of 584 amino acids, molecular weight 66.9 kDa, and isoelectric point 5.79. The nucleotide and amino acid sequences of Nanglai-BmNPV *ie-1* gene showed, respectively, 98 and 97% identity to those of the BmNPV *ie-1* genes. The identity of the Nanglai-BmNPV IE-1 deduced amino acid sequence with IE-1 protein sequences of other 37 known Lepidopteran NPVs varied from 20% up to 97%. The phylogenic tree established with the IE-1 protein sequences of the Lepidopteran NPVs showed similar topology to those previously reported using highly conserved genes. Interestingly, the tree obtained with IE-1 clearly shows that the Lepidopteran NPVs group II clade can be subdivided into three subgroups.

KEYWORDS: silkworm, viral propagation, silkworm disease

INTRODUCTION

Bombyx mori nucleopolyhedrovirus (BmNPV) is the most devastating pathogen causing grasserie, the worst silkworm deleterious disease. BmNPV is a member of the genus *Nucleopolyhedrovirus* (NPV), family Baculoviridae. The NPVs of this family have been reported to be pathogens infecting over 600 species of arthropods, particularly insects of the order Lepidoptera[1]. The genomes of 37 Lepidopteran NPVs including BmNPV[2] have been sequenced completely and are available in the NCBI database. The Lepidopteran NPVs were first separated into group I and group II, based on the *polyhedrine* gene[3]. Such division into two groups has been supported by phylogenetic analysis based on other genes[4–7].

In the infection cycle, NPVs produce two viral types, budded virus (BV) and occlusion-derived virus or occluded virus (ODV) for effective viral propagation. The BVs are produced in the nucleus of infected cells and are released to infect neighbouring cells within an infected insect. In the last phase of infection, the viruses are encapsulated in a protein envelope becoming ODVs, which are released to the environment and transmitted to other insects via oral infection[8]. Genes involved in the proliferation of NPVs are divided into three classes, early (immediate-early and delayed early), late, and very late genes, based on their expression time after virus infection. These genes are expressed coordinately in a sequential order[9,10]. In the early phase of viral infection, the early genes are expressed in cascade since they are required for viral DNA replication and regulation of gene expression. The late genes are expressed after 6–8 h postinfection and required for the BV production, whereas the very late genes required to produce ODV are expressed beyond 12 h postinfection[11].

The immediate early-1 (*ie-1*) gene is one of the seven essential genes required for viral DNA replication[9,12–14]. The gene is expressed immediately after viral infection and its product, IE-1, functions as a regulatory protein for the transcription of other genes[15–18]. Thus the *ie-1* gene may be a key gene in the regulation of the viral gene expression cascade. In this study, the *ie-1* gene of nucleopolyhedrovirus infecting a Thai silkworm, Nanglai, was identified

and compared with all known NPV *ie-1* genes of Lepidopteran NPVs available in the database. Since BmNPV causes dramatic loss in Thai silk industry, the information about the Nanglai-BmNPV *ie-1* gene may be useful in suppressing of NPV proliferation and eventually eliminating the most serious silkworm disease, grasserie. In addition, the phylogenetic relationship of Lepidopteran NPVs based on the deduced amino acid sequences of the IE-1 are analysed and discussed.

MATERIALS AND METHODS

NPVs purification and DNA extraction

BmNPVs were purified from infected larvae of a Thai silkworm, Nanglai, collected from a farm in Srisaket province, Thailand. Genomic DNA of Nanglai-BmNPV was extracted by a method modified from that reported in Chaeychomsri[19]. Briefly, the polyhedral inclusion bodies were extracted from the haemolymph of infected larvae using a refrigerated ultracentrifuge. The alkaline solution (0.2 M Na_2CO_3, 0.5 M EDTA, and 0.34 M NaCl) was added at the ratio of 1:1 and incubated for 30 min on ice prior to clear solution collection by centrifugation at 14 000g for 1 min. Then, 500 µl of 2% SDS, 10 µl of Proteinase K (10 mg/ml), and 8 µl of RNaseA (10 mg/ml) were added to 500 µl of the clear solution. The mixture was incubated at 37 °C for 1 h and then centrifuged at 14 000g for 5 min. The supernatant was collected for extraction of the viral genomic DNA by using the phenol/chloroform method. The DNA pellet was dissolved in TE buffer and kept at −20 °C.

Cloning and sequencing

Two sets of primers were designed using the sequence of BmNPV T3 (NC 001962) available in the NCBI database. Primer set 1 was F: 5′-AGCGGGTGGCGGCTCG-3′ and R: 5′-CACATTGCTCACGTAATTGGC-3′ yielding the 1454 bp PCR product and primer set 2 was F: 5′-ATGTCCGAGACTGGTTATTAC-3′ and R: 5′-GCTAACGATTCGCCCCTACT-3′ giving the 1624 bp PCR product. PCR conditions were as follows: preheated at 94 °C for 5 min, followed by 35 cycles of denaturing at 95 °C for 1 min, annealing at 64 °C for 1 min and extending at 72 °C for 3 min. PCR products were detected by 1% agarose gel. The products were cloned into pGEM-T Easy and then sent to Macrogen, Korea, for sequencing.

Sequence alignment and phylogenetic analysis

The sequences obtained from the two PCR products were assembled. Both the nucleotide and the deduced amino acid sequences of Nanglai-BmNPV *ie-1* gene were then aligned with those of 37 other known Lepidopteran NPVs whose genomes have been completely sequenced, using the CLUSTALW version 1.81. Phylogenetic trees based on IE-1 protein sequences were constructed using neighbour joining and maximal parsimony methods, P distance model, MEGA version 4.0[20]. The reliability of clusters within the tree was evaluated based on 1000 bootstrap replications.

RESULTS AND DISCUSSION

Nucleotide and deduced amino acid sequence

The combined PCR fragment obtained in this study was 2942 bp long and covered the whole region of *ie-1* gene, the upstream flanking genes, open reading frame (ORF) 122, and a part of the downstream flanking gene, *odv-e 56*. The ORF of Nanglai-BmNPV *ie-1* gene comprised 1755 nucleotides encoding a deduced protein of 584 amino acids (Fig. 1). The expected molecular weight of the protein was 66.9 kD, as calculated by the pI/MW tool of Expasy web (http://www.expasy.ch/tools/pi-tool.html). The Nanglai-BmNPV *ie-1* ORF sequence was rich in AT, with 63.2% AT and 36.8% GC. The transcription start site (+1)[21] was located 50 bp upstream of the translation initiation codon, ATG. The termination codon, TAA, was located at the position +1802 to +1804 followed by a 3′ untranslated region containing a putative polyadenylation site (AATAAA) at the position +2006 to +2011 (Fig. 1).

The nucleotide and deduced amino acid sequences of Nanglai-BmNPV *ie-1* were compared with those of two NPVs infecting *Bombyx mori*, BmNPV T3[2] and BmNPV K1[22], available in the NCBI database using CLUSTALW. Both BmNPV T3 and BmNPV K1 were infected silkworms of bivoltine race, while Nanglai-BmNPV infected silkworms of polyvoltine race. The *ie-1* ORFs and IE-1 proteins of these three NPVs were of equal length, 1755 bp and 584 amino acid residues.

The identity of nucleotide and amino acid sequences of Nanglai-BmNPV *ie-1* gene to those of BmNPV T3 and BmNPV K1 were 98% and 97%, respectively. The *ie-1* ORF of Nanglai-BmNPV differed from that of the BmNPV T3 at 26 positions: transition substitution type occurred at 19 positions, with 12 G ⟷ C and 7 T ⟷ C, while the transversion substitution type occurred at 7 positions. The deduced amino acid sequence of Nanglai-BmNPV IE-1 and that of BmNPV T3 differed at 12 positions. When compared with BmNPV K1 *ie-1* gene, the nucleotide sequence of Nanglai-BmNPV *ie-1* showed 31 posi-

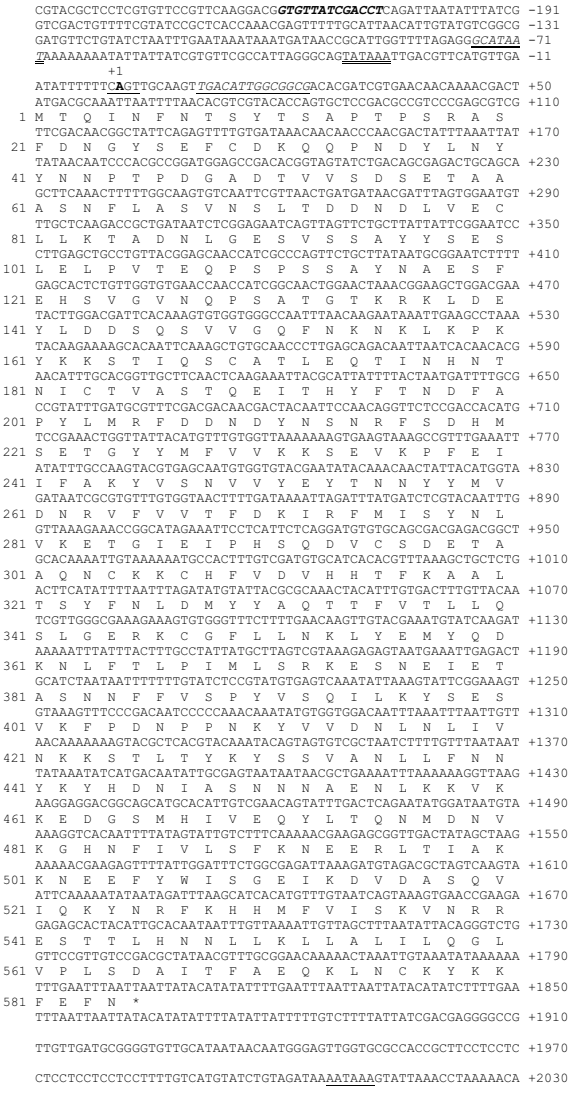

Fig. 1 Nucleotide sequence, deduced amino acid, and the 5′ noncoding sequences of Nanglai-BmNPV *ie-1* gene. The numbers with a plus sign indicate the nucleotide position from the transcription initiation site (+1). The amino acid sequence is numbered sequentially from the first amino acid, the methionine. In the 5′ noncoding region, the ecdyson response element is bolded and in italics, the CAAT-like motif is in italics and double underlined, the putative TATA box is double underlined, the early promoter is underlined and the DAR sequence is underlined and in italics. The polyadenylation signal sites are underlined.

tions of nucleotide substitution, 23 of which were transition substitution type (15 G $\longleftrightarrow$ A, and 8 C $\longleftrightarrow$ T), and the remaining 8 substitutions were transversion type. Differences between IE-1 sequences of Nanglai-BmNPV and BmNPV K1 were found at 15 positions. The nucleotide substitutions often occurred in the 5′ region of the *ie-1* nucleotide sequence, which corresponds to the N-terminal of the amino acid sequence. The Nanglai-BmNPV IE-1 deduced amino acid sequence was further aligned with other 37 deduced IE-1 protein sequences of NPVs whose genomes have been completely sequenced and are available in the NCBI database using CLUSTALW 1.81. The result showed that the identity of Nanglai-BmNPV IE-1 to the 37 IE-1 proteins ranged from 20% up to 97%. The highly variable identity among IE-1 proteins indicated that Nanglai-BmNPV *ie-1* gene is a poorly conserved gene.

Analysis of 5′ noncoding sequence

The 5′ noncoding region of Nanglai-BmNPV *ie-1* gene was searched for conserved regions within 300 bp upstream from the translation start site (ATG). Five conserved sequences were found in this region. They were a downstream activating region (DAR, TGACATTGGCGGCG), the tetranucleotide, CAGT, motif, a TATA (TATAAA) element, a putative CAAT box (GCATAAT), and an ecdysone response element (GTGTTATCGACCT). The CAGT motif functioning as an early promoter was located at the position -1 to $+3$ of 51 bp upstream of the translation start site, ATG. The TATA element was located at the position -26 to -31, 31 bp upstream of the transcription start site ($+1$). The location of the DAR was at position $+11$ to $+24$ and 10 bp downstream of the transcription start site. The CAAT box was found at position -70 to -76, whereas the ecdysone response element was located at -208 to -220 (Fig. 1).

The 300 bp 5′ noncoding sequence upstream of the Nanglai-BmNPV *ie-1* translation start site (ATG) was compared with other 37 Lepidopteran NPVs available from the NCBI database (Table 1). The result showed that the *ie-1* genes from 7 NPVs, namely AcNPV, BmaNPV, BmNPV T3, Nanglai-BmNPV, MvMNPV, PxMNPV, and RoMPNV, contained the five conserved identical sequences in their 5′ noncoding regions. Interestingly, the distance between the TATA element and CAGT motif and between the CAGT motif and the translation start site (ATG) in these seven NPVs were 31 bp and 51 bp, respectively, in all 7 *ie-1* promoters. The locations of DAR sequence in these 7 promoters were at nucleotide $+11$ to $+24$ as previously reported[23]. The TATA element and CAGT motif are often found in the promoter region of immediately early genes of baculovirus[2,24,25]. It has been suggested that the CAGT motif plays an important role as an initiator element for the regulation of *ie-1* transcription since substitution of a

Table 1 List of Lepidopteran NPVs used in this study. Their genomes have been completely sequenced and available in the NCBI database.

Species	Abbreviation	Accession	Amino acid
Adoxophyes honmai NPV	AhNPV	NC_004690	642
Adoxophyes orana NPV	AoNPV	NC_011423	645
Antheraea pernyi NPV	AnpeNPV	NC_008035	577
Anticarsia gemmatalis MNPV	AgMNPV	NC_008520	588
Agrotis ipsilon multiple MNPV	AgipMNPV	NC_011345	675
Agrotis segetum NPV	AgseNPV	NC_007921	661
Autographa californica MNPV	AcMNPV	NC_001623	636
Bombyx mandarina NPV	BmaNPV	NC_012672	584
Bombyx mori NPV	BmNPV	NC_001962	584
Choristoneura fumiferana DEF MNPV	CfDEFMNPV	NC_005137	560
Choristoneura fumiferana MNPV	CfMNPV	NC_004778	565
Chrysodeixis chalcites NPV	CcNPV	NC_007151	742
Clanis bilineata NPV	CbNPV	NC_008293	722
Ecotropis obliqua NPV	EoNPV	NC_008586	721
Epiphyas postvittana MNPV	EppoMNPV	NC_003083	556
Euproctis pseudoconspersa NPV	EupsNPV	NC_012639	612
Helicoverpa armigera NPV	HaSNPV	NC_003094	661
Helicoverpa armigera MNPV	HaMNPV	NC_011615	601
Helicoverpa armigera NPV G4	HaSNPV-G4	NC_002654	655
Helicoverpa armigera NPV NNg1	HaSNPV NNg1	NC_011354	655
Helicoverpa zea SNPV	HzSNPV	NC_003349	655
Hyphantria cunea NPV	HycNPV	NC_007767	560
Lymantria dispar MNPV	LydMNPV	NC_001973	566
Lymantria xylina MNPV	LyxMNPV	NC_013953	569
Leucania separata NPV	LesNPV	NC_008348	927
Mamestra configurata NPV-A	McNPV-A	NC_003529	526
Mamestra configurata NPV-B	McNPV-B	NC_004117	603
Maruca vitrata MNPV	MvMNPV	NC_008725	575
Orgyia leucostigma NPV	OlNPV	NC_010276	560
Orgyia pseudotsugata MNPV	OpMNPV	NC_001875	639
Plutella xylostella MNPV	PxMNPV	NC_008349	582
Rachiplusia ou MNPV	RoMNPV	NC_004323	580
Spodoptera exigua MNPV	SeMNPV	NC_002169	714
Spodoptera frugiperda MNPV	SfMNPV	NC_009011	382
Spodoptera litura NPV	SlNPV	NC_003102	688
Spodoptera litura NPV II	SlNPV II	NC_011616	707
Trichoplusia ni single SNPV	TnSNPV	NC_007383	733

single nucleotide in the CAGT motif reduces the *ie-1* RNA level[25]. However, it is still unknown how the CAGT motif regulates the transcription of baculovirus genes[26]. Of 37 *ie-1* genes, 3 from AhNPV, AoNPV, and EoNPV, had neither TATA nor CAGT motif in their 5′ noncoding regions. It has been suggested that the transcription of the *ie-1* gene having no early promoter might be regulated by a distal promoter[27]. Moreover, the early promoters of HaSNPV, HaSNPV-NNgI, HaSNPV-G4, and HzSNPV, contained TATA sequences and CATT motifs instead of CAGT motifs[25]. Excluding the above 7 NPVs, the remaining 31 NPVs did not contain identical sequences to the DAR and the ecdysone response element of Nanglai-BmNPV. In AcNPV, the DAR functioned as a stimulator of *ie-1* transcription increasing 10-fold the *ie-1* RNA level during infection. However, *ie-1* is not transcribed if both CAGT motif and DAR sequences are simultaneously deleted[23].

The ecdysone response element in the promoter of BmNPV *ie-1* gene was identified by a series of truncated *ie-1* promoters in BmN cells. It has been suggested that the activity of *ie-1* regulated by ecdysone might resist viral replication[28]. Ecdysone is a molting hormone in insects. The molting of an insect infected with NPV would be obstructed by the function of the viral gene, *ecdysteroid UDP-glucosyltransferase* (*egt*). The *egt* gene encodes enzyme EGT catalyses the inactivation of ecdysone hormones of a host insect, resulting in an extension of the larval period, since molting and pupation require ecdysones[29]. The extension of infected larval period provides more multiplication time resulting in an increased viral number. In addition, in infected insects, the inactive ecdysone may not bind the ecdysone response element of *ie-1* promoter so the activity of *ie-1* to promote viral replication may also be inhibited.

Phylogenetic analysis of group I

Phylogenetic tree of Lepidopteran NPVs was constructed with the deduced IE-1 amino acid sequences from Nanglai-BmNPV and with 37 NPVs available from the NCBI database using the neighbour-joining (NJ) and maximal parsimony (MP) methods, P distance model, MEGA version 4[20]. The IE-1 (ORF9) of *Xestia c-nigrum* granulovirus (XcnGV, NP_059157.1) was used as an outgroup. The trees obtained from both methods had very similar topologies. However, the bootstrap values of the NJ tree were higher than those of the MP tree in all branches (Fig. 2). Both trees clearly showed that Lepidopteran NPVs were divided into two major clades, group I clade and group II clade, consistent with all previous reports. Group I clade was also separated into two subclades, clade I-A and clade I-B. As expected, the trees grouped Nanglai-BmNPV into clade I-A (Fig. 2). The members of clade I-A were 7 NPVs, namely Nanglai-BmNPV, BmNPV, BmaNPV, RoMNPV, PxMNPV, AcMNPV, and MvMNPV. The Nanglai-BmNPV was closely related to the two NPVs infecting *Bombyx* spp., BmNPV and BmaNPV. In clade I-B, seven NPVs, CfDEFMNPV, AgMNPV, EppoMNPV, OpMNPV, CfMNPV, HycNPV, and AnpeNPV were clustered together. However, the topologies of NJ tree (Fig. 2A) and MP tree (Fig. 2B) were slightly different. Like most previously reported trees, the IE-1 tree topology placed AcMNPV within group I-A. In contrast, the tree topology generated by the *polyhedron* (*polh*) gene placed AcMNPV in group II and it was suggested that the *polh* gene of AcMNPV might have recombined with that of a virus in group II[30–32]. The Lepidopteran NPVs were separated into two groups, group I and II, based on the conserved gene, polyhedron, which encoded the major occlusion body protein required for producing viral progeny[3,5]. Later, the phylogenetics of NPVs were generated with other conserved genes,

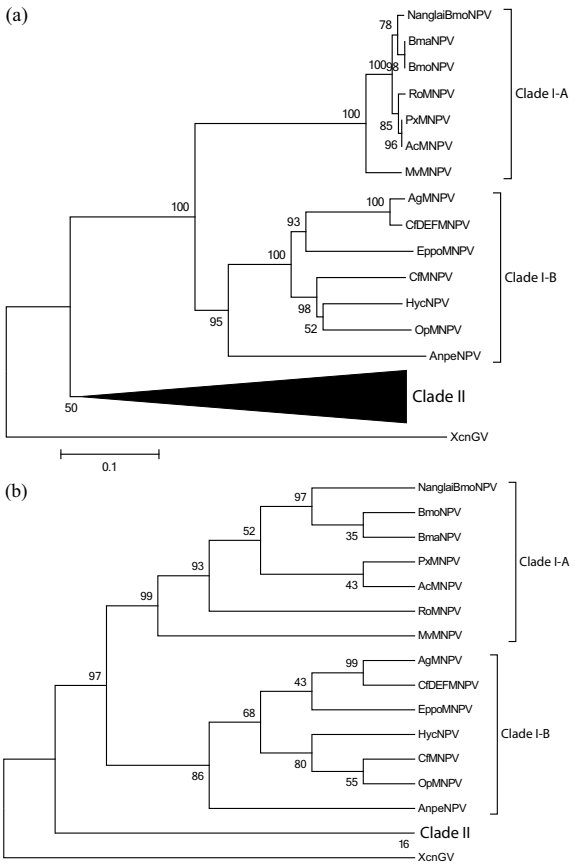

Fig. 2 Phylogenetic trees based on the deduced amino acid sequence of IE-1 proteins of Lepidopteran NPV group I clade generated by (a) neighbour joining and (b) maximal parsimony methods. Numbers at the nodes indicate the values from 1000 bootstrap replications. *Xestia c-nigrum* granulovirus (XcnGV, NP_059157.1) was used as an outgroup. In both trees, NPVs of group I clade was separated into two subclades, subclade I-A and subclade I-B.

for example *lef-2*[33], *lef-8*, *lef-9*[34,35], and *ecdysteroid UBP-glycosyl transferase* (*egt*)[4]. Recently, many phylogenetics of NPVs have been established by means of the combination of several conserved gene sequences and whole genome sequences[7,36,37]. These phylogenetic trees were suggested to be more acceptable than those constructed by a single gene. However, the tree constructed by a single gene had been used for quick identification[34]. In this work, even though *ie-1* gene was poorly conserved, the pattern of the tree topology derived from IE-1 amino acid sequences agreed with those of other tree topologies based on single highly conserved genes, the combination of highly conserved genes, and complete genome sequences.

Interestingly, the 5′ noncoding sequences of the *ie-1* genes from group I-A NPVs were highly conserved: all of them contained the TATA element, CAGT motif, CAAT-like motif, DAR region, and ecdysone response sequence. Moreover, the interval distances between the TATA element and CAGT motif (31 bp long) and between the CAGT motif and the translation start site, ATG, (51 bp long) were identical. Therefore, the tree topology generated using the 5′ noncoding sequence (300 bp upstream of ATG), showed that the Lepidopteran NPVs group I was divided into two subclades, A and B (data not shown). Moreover, the result of group I NPV IE-1 sequence alignment showed that the IE-1 sequences of the three NPVs, BmaNPV, BmNPV, and Nanglai-BmNPV, infecting *Bombyx* spp. insects, contained unique amino acids at 12 positions differing from those present in the other NPVs (Fig. 3). Eight of these 12 conserved amino acids were located in the N-terminal part of the IE-1 protein. The N-terminal region of the IE-1 protein contains an acidic domain that is essential for the viral transactivational activity[37–40]. The remaining 4 unique amino acid positions were located in the C-terminal region of IE-1. The C-terminal region contains a domain required for inhibitory and DNA-binding activities[37]. Only one of these 4 unique amino acid positions was in a putative single-stranded-DNA-binding motif that may be involved in DNA binding[12]. It is likely that the 12 unique amino acid positions of IE-1 proteins from *Bombyx* spp. NPVs may provide these NPVs with the specific ability to infect only *Bombyx* spp. insects.

Phylogenetic analysis of group II

The remaining 24 NPVs were grouped into group II clade. In most previous reports, the division of NPV group I into two subgroups was quite clear, but the taxon of NPVs within group II was inconsistent. Even if a number of complete genome sequences available to date were used to construct a phylogenetic tree, the division within group II was not well defined. It was suggested that the relationship of NPVs within group II was more diverse than that of group I clade[6,41].

In this study, trees based on IE-1 sequences from the 24 NPVs of group II clade were generated using the NJ and MP methods. Both methods established trees with very similar topology (Fig. 4). In both trees, three clear subclades within the group II clade were derived from a monophyletic origin. The names of the three subclades, designated A, B, and C, were given following the previous report in which the tree was constructed with the DNA polymerase amino acid sequences of 17 NPVs[5]. Of the 17 NPVs, only 7

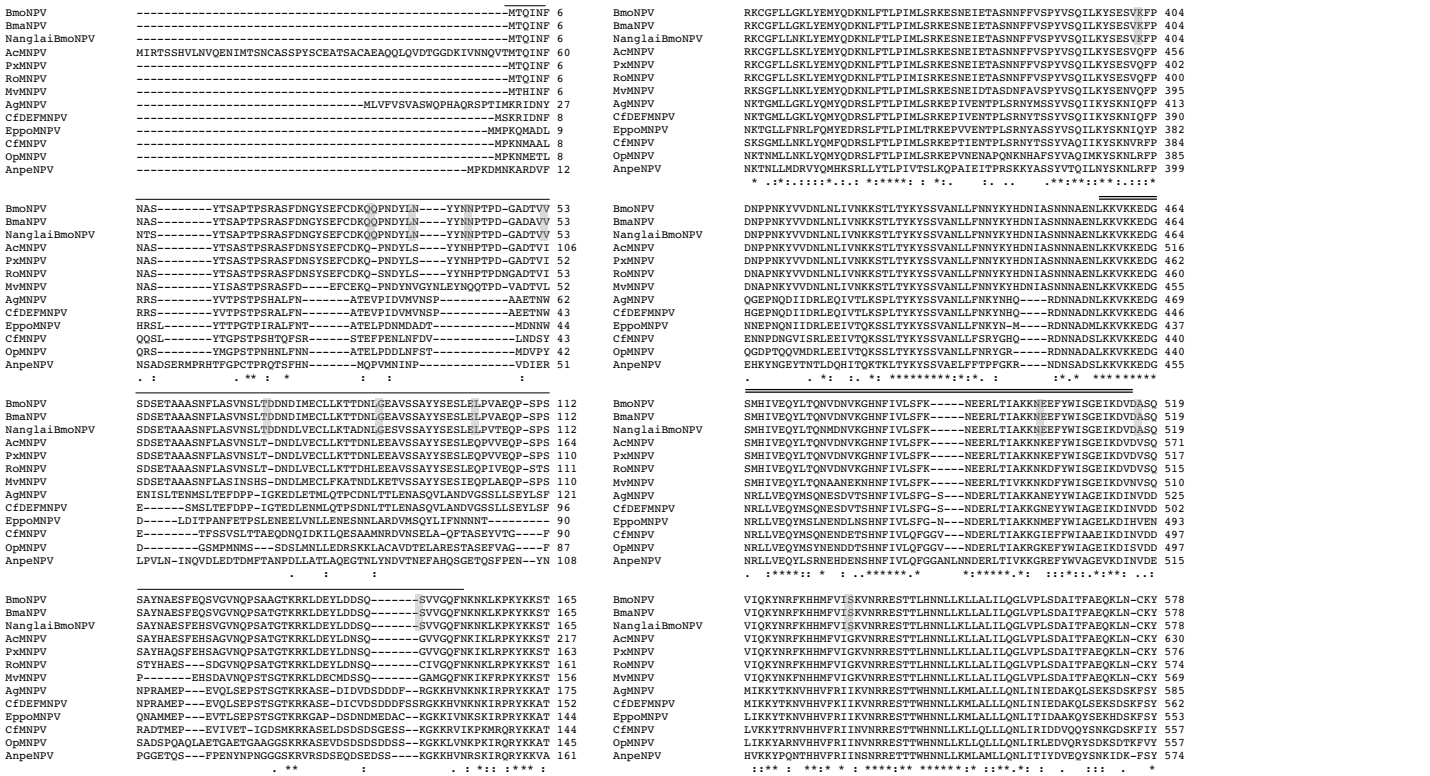

Fig. 3 Multiple alignment of the IE-1 deduced amino acid sequences of the group I clade NPVs using CLUSTALW version 1.81. Labelled boxes indicated unique amino acids at 12 positions of the IE-1 protein from NPVs infecting *Bombyx* spp. insects. The N-terminal transactivation domain was marked by a line over the columns. The SSB-like motif was indicated by double line over the columns.

NPVs had IE-1 sequences available on the database and were used in this study. In the previous report, of the 7 NPVs, SeMNPV and SfMNPV were in subgroup II-A; SlNPV, HaSNPV, and HzSNPV were in subgroup II-B; and EoNPV was in subgroup II-C. LydMNPV was not grouped into any of the three subgroups. In this study, subclade II-A was composed of 10 NPVs, namely McNPV-A, McNPV-B, HaMNPV, TnSNPV, CcNPV, AgseNPV, AgipMNPV, SfMNPV, SlNPV II, and SeMNPV. Subclade II-B contained 6 NPVs, HaSNPV, HaSNPV-NNg1, HaSNPV-G4, HzSNPV, SlNPV, and LesNPV. The last one, subclade II-C, consisted of 8 NPVs, OlNPV, EoNPV, EupsNPV, CbNPV, LyxMNPV, LydMNPV, AoNPV, and AhNPV. The reliability of these subclades within the group II clade was supported by 1000 bootstrap resampling value. In the NJ tree, the bootstrap value at the root node of the three subclades was low, only 50%. However, the bootstrap gave strong support to the existence of subclade II-A (79%) and II-B (76%), while subclade II-C was weakly supported by the bootstrap value (42%). The outcome of the MP tree agreed with that of the NJ tree, even though the bootstrap values supporting the existence of the three subclades within group II clade NPVs were lower than those of the NJ tree. For example, the existence of subclade II-A and II-B in the MP tree was supported by 57% and 76% bootstrap values, respectively, while subclade II-C was supported by a low bootstrap value. Nevertheless, in both trees, the members within each subclade were identical. The difference between the NJ tree and the MP tree was the position of CbNPV. In the NJ tree, CbNPV was a sister group of OlNPV, EoNPV, and EupsNPV confirmed by 71% bootstrap value (Fig. 4), whereas in the MP tree, CbNPV was a sister group of LyxMNPV, LydMNPV, AoNPV, and AhNPV with a very low bootstrap value (Fig. 4B).

In this study, the identity percentages of IE-1 sequences were determined by pairwise comparisons. The identity percentages obtained from alignment of 38 Lepidopteran NPV IE-1 deduced amino acid sequences were used to calculate the average identity percentages of each subclade and between subclades, by CLUSTALW (Table 2). From Table 2, the average identity of IE-1 sequences within subclade I-A was the highest, 93%, whereas the lowest average

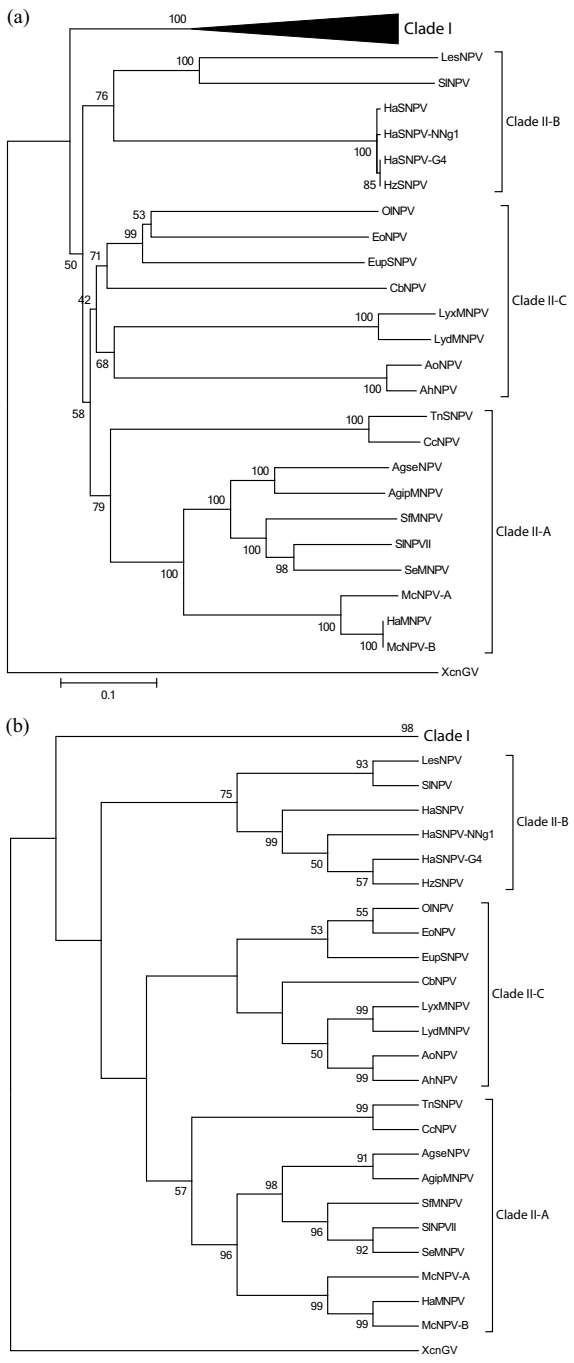

Fig. 4 Phylogenetic trees present three subclades, A, B, and C, within Lepidopteran NPV group II clade based on the deduced amino acid sequence of IE-1 proteins generated by (a) neighbour joining and (b) maximal parsimony methods. Bootstrap values (1000 replications) lower than 40% were not indicated at the nodes. *Xestia c-nigrum* granulovirus (XcnGV, NP_059157.1) was used as an out group.

Table 2 Pairwise comparison within subclades and between subclades of IE-1 deduced amino acid sequences of Lepidopteran NPVs.

Pairwise Comparison	Average identity (%)		Max. identity	Min. identity
	Mean	SD		
Within clade I-A	93.1	4.4	99	86
Within clade I-B	66.	11.	96	44
Within clade II-A	44.	18.	99	24
Within clade II-B	56.	37.	100	25
Within clade II-C	33.	15.	88	24
Clade I-A > Clade I-B	44.7	4.1	50	34
Clade I-A > Clade II-A	24.0	1.8	28	20
Clade I-A > Clade II-B	23.0	2.1	26	17
Clade I-A > Clade II-C	23.4	2.4	29	18
Clade I-B > Clade II-A	24.0	2.1	28	19
Clade I-B > Clade II-B	22.6	2.2	26	18
Clade I-B > Clade II-C	22.7	1.7	27	20
Clade II-A > Clade II-B	26.4	3.4	39	19
Clade II-A > Clade II-C	26.8	3.6	35	20
Clade II-B > Clade II-C	25.0	3.3	31	19

identity, 33%, was found in subclade II-C. Thus the IE-1 sequences belonging to NPVs within clade I-A were less divergent than those of other subclades. Low identity percentages were obtained in pairwise comparisons between subclades (Table 2). These results showed clear divergence of the IE-1 sequences between any subclade of group I and group II, and between subclades within group II. Even though IE-1 sequence has diverged more in Lepidopteran NPVs compared to other genes, the amino acid sequence consistently identified the taxons within Lepidopteran NPVs, particularly within the group II clade.

Acknowledgements: We gratefully thank Dr Nora Facessti for her suggestion and English proofreading. This work was granted from Kasetsart University Development Institute and Faculty of Science, KU, and Centre for Excellence in Silk, Kasetsart University, for LN and Kasetsart Graduate School for BCh.

REFERENCES

1. Martignoni ME, Iwai PJ (1981) A catalogue of viral diseases of insects, mites and ticks. In: Burges HD (ed) Microbial control of pests and plant diseases 1970 - 1980, Academic Press, London, pp 897–911.
2. Gomi S, Majima K, Maeda S (1999) Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *J Gen Virol* **80**, 1323–37.
3. Zanotto PMA, Kessing BD, Maruniak JE (1993) Phylogentic interrelationships among baculoviruses: evo-

lutionary rates and host associations. *J Invertebr Pathol* **62**, 147–64.
4. Chen XW, Hu ZH, Jehle JA, Zhang YQ, Vlak JM (1997) Analysis of the *ecdysteriod UDP-glucotransferase* gene of *Heliothis armigera* single nucleocapsid baculovirus. *Virus Genes* **15**, 219–25.
5. Bulach DM, Kumar CA, Zaia A, Liang B, Tribe ED (1999) Group II nucleopolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedron and DNA polymerase gene sequences. *J Invertebr Pathol* **73**, 59–73.
6. Herniou AE, Luque T, Chen X, Vlak MJ, Winstanley D, Cory SJ, O'Reilly RD (2001) Use of whole genome sequence data to infer Baculovirus phylogeny. *J Virol* **75**, 8117–26.
7. Mukawa S, Goto C (2006) In vivo characterization of a group II nucleopolyhedrovirus isolated from Mamestra brassicae (Lepidoptera: Noctuidae) in Japan. *J Gen Virol* **87**, 1491–500.
8. Egelhard EK, Kam-Morgan LN, Washburn JO, Volkman LE (1994) The insect tracheal system: a conduit for the systemic spread of *Autographa california* nuclear polyhedrosis virus. *Proc Natl Acad Sci USA* **91**, 3224–7.
9. Lu A, Miller L (1995) The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *J Virol* **69**, 975–82.
10. Yamagishi J, Isobe R, Takebuchi T, Bando H (2003) DNA microarrays of baculovirus genomes: differential expression of viral genes in two susceptible insect cell lines. *Arch Virol* **148**, 587–97.
11. Blissard G, Rohrmann GF (1990) Baculovirus diversity and molecular biology. *Ann Rev Entomology* **35**, 127–55.
12. Kool M, Ahrens CH, Goldbach RW, Rohrmann GF, Vlak JM (1994) Identification of genes involved in DNA replication of the *Autographa californica* baculovirus. *Proc Natl Acad Sci USA* **91**, 11212–6.
13. Lu A, Krell O, Vlak J, Rohrmann G (1997) The Baculoviruses. Plenum Press, New York, pp 193–216.
14. Lin G, Blissard GW (2002) Analysis of an *Autographa californica* nucleopolyhedrovirus lef-11 knockout: LEF-11 is essential for viral DNA replication. *J Virol* **76**, 2770–9.
15. Guarino LA, Summer MD (1988) Functional mapping of *Autographa californa* nuclear polyhedrosis virus genes required for late gene expression. *J Virol* **62**, 3924–8.
16. Carson DD, Summer MD, Guarino LA (1991) Molecular analysis of a baculovirus regulatory gene. *Virology* **182**, 279–86.
17. Passareilli A, Miller L (1993) Three baculovirus genes involved in late and very late gene expression: *ie-1*, *ie-n*, and *lef-2*. *J Virol* **67**, 2149–58.
18. Olson VA, Wetter JA, Friesen PD (2002) Baculovirus transregulator IE1 requires a dimeric nuclear localization element for nuclear import and promoter activation. *J Virol* **76**, 9505–15.
19. Chaeychomsri S (2003) Purification of virus and extraction of viral DNA. pp 1–8. In A Laboratory Manual of a Regional Workshop on Molecular Biology and Applications of baculovirus. Central Laboratory and Greenhouse Complex Kasetsart Univ, Nakhon Pathom, Thailand.
20. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–9.
21. Kojima K, Hayakawa T, Asano S, Bando H (2001) Tandem repetition of baculovirus ie-1 promoter results in upregulation of transcription. *Arch Virol* **146**, 1407–14.
22. Park HJ, Lee KS, Je YH, Sohn HD, Jin BR (2001) Molecular cloning and sequence analysis of the immediate early viral gene, IE-1, from *Bombyx mori* nuclear polyhedrosis virus K1. *Int J Indust Entomol* **3**, 43–9.
23. Pullen S, Friesen PD (1995) Early transcription of the ie-1 tranregulator gene of *Autographa California* nuclear polyhedrosis virus is mediated by DNA sequence within 5′ noncoding leader. *J Viol* **69**, 156–65.
24. Pullen S, Friesen PD (1995) The CAGT motif functions as an initiator element during early transcription of the Baculovirus transregulator *ie-1*. *J Virol* **69**, 3575–83.
25. Xing K, Deng R, Wang J, Feng J, Huang M, Wang X (2005) Analysis and prediction of baculovirus promoter sequences. *Virus Res* **113**, 64–71.
26. Mikhaailov VS (2003) Replication of the Baculovirus genome. *Molecular Biology* **37**, 250–9.
27. Hilton S, Winstanley D (2008) Genomic sequence and biological characterization of a nucleopolyhedrovirus isolated from the summer fruit tortrix, *Adoxophyes orana*. *J Gen Virol* **89**, 2898–08.
28. Kojima K, Oritani K, Nakatsukasa T, Asano S, Sahara K, Bando H (2007) Ecdysone response element in a baculovirus immediate-early gene, *ie-1*, promoter. *Virus Res* **130**, 202–9.
29. O'Reily DR, Miler LK (1991) Improvement of a baculovirus pesticide by deletion of the *egt* gene. *Biotechnology* **9**, 1086–9.
30. Clarke EE, Tristem M, Cory J, O'Reilly RD (1996) Characterization of the *ecdysteroid UDP-glucosyltransferase* gene from *Mamestra brassica* nuleopolyhedrovirus. *J Gen Virol* **77**, 2865–71.
31. Harrison RL, Bonning CB (2003) Comparative analysis of the genome of *Rachiplusia ou* and *Autographa californica* multiple nucleopolyhedroviruses. *J Gen Virol* **84**, 1827–42.
32. Jehle JA (2004) The mosaic structure of the polyhedrin gene of the *Autographa californica* nucleopolyhedrovirus (AcMNPV). *Virus Genes* **29**, 5–8.
33. Chen X, Ijkel WFJ, Dominy C, Zanotto P, Hashimoto Y (1999) Identification, sequence analysis and phylogeny of *lef-2* gene of *Helocoverpa armigera* single-nucleocapsid baculovirus. *Virus Res* **65**, 21–32.
34. Lange M, Wang H, Zhihong H, Jehle JA (2004) Towards a molecular identification and classification

system of lepidopteran-specific baculoviruses. *Virology* **325**, 36–47.
35. Jehle JA, Lange M, Wang H, Hu Z, Wang Y, Hauschild R (2006) Molecular identification and phylogenetic analysis of baculoviruses from Lepidotera. *Virology* **346**, 180–93.
36. Oliveira JVC, Wolff JLC, Garcia-Maruniak A, Ribeiro BM, de Castro MEB, de Souza ML, Moscardi F, Maruniak JE, Zanotto PMA (2006) Genome of the most widely used viral biopesticide: *Anticarsia gemmetalis* multiple nucleopolyhedrovirus. *J Gen Virol* **87**, 3233–50.
37. Nai YS, Wu CY, Wang TC, Chen YR, Lau WH, Lo CF, Tsai MF, Wang CH (2010) Genomic sequencing and analyses of *Lymantria xylina* multiple nucleopolyhedrovirus. *BMC Genom* **11**, 116.
38. Kovacs GR, Choi J, Guarino LA, Summers MD (1992) Functional dissection of the *Autograph californica* nuclear polyhedrosis virus immediate-early 1 transcriptional regulatory protein. *J Virol* **66**, 7429–37.
39. Rodems SM, Pullen SS, Friesen PD (1997) DNA-dependent transregulation by IE-1 of *Autographa californica* nuclear polyhedrosis virus and DNS binding. *J of Virol* **71**, 9270–7.
40. Slack MJ, Blissard WG (1997) Identification of two independent transcriptional activation domains in the *Autographa californica* multicapsid nuclear polyhedrosis virus IE-1 protein. *J of Virol* **71**, 9579–87.
41. Lange M, Jehle JA (2003) The genome of the *Cryptophlebia leucotreta* granulovirus. *Virology* **317**, 220–36.