Int. J. Wild Silkmoth & Silk **12**, 1–8 (2007)

Nucleotide Sequence Analysis of *Pst*I D Fragment of *Antheraea pernyi* Nucleopolyhedrovirus Clone A and Identification of *Escherichia coli* Insertion Sequence

Yuan Jiao Huang[1], Hiroshi Mitsutake[2], Makoto Itakura[2], Xue Ying Wang[3] and Jun Kobayashi[2,*]

[1] *Guangxi Medical University, Nanning 530021, China*

[2] *Faculty of Agriculture, Yamaguchi University, Yamaguchi 753-8515, Japan*

[3] *Shenyang Agricultural University, Shenyang 110161, China*

* *To whom correspondence should be addressed. E-mail : koba-jun@yamaguchi-u.ac.jp*

Abstract **The nucleotide sequence of *Pst*I D fragment (6318 bp) of a plaque-purified clone (clone A) of the *Antheraea pernyi* nucleopolyhedrovirus (AnpeNPV) was determined and compared with recently published complete genome sequence of another AnpeNPV isolate (Liaoning strain) (GenBank accession number NC_008035). Eleven open reading frames (ORFs) were identified in the fragment and completely or almost identical to ORFs (from Anpe101 to Anpe111) of the Liaoning strain, except for Anpe105. A putative protein encoded by Anpe105 is truncated in the clone A by a 10 bp-deletion within the coding region, which causes a frame-shift and the earlier translation termination. Further comparative sequence analysis among baculoviruses has revealed that several group I NPVs possess Anpe105 homologs. All of these homologs encode much larger proteins with a conserved domain that Anpe105 product lacks. These results suggested that Anpe105 might be nonessential or functionless for the AnpeNPV replication cycle and, therefore, the genome sequence of Anpe105 region is unstable and tends to get deletions and/or insertions, resulting in the sequence heterogeneity among AnpeNPV isolates. In addition, we obtained two mutants of *Pst*I D fragment cloned in the plasmid during the propagation in bacterial cells. Sequence analysis revealed that one of the mutants had an insertion of IS10 transposon (1329 bp) between ORFs corresponding to Anpe107 and Anpe108, and the other mutant had a large deletion (4912 bp) corresponding to a region from Anpe101 to Anpe110. Implications of these mutants are discussed.**

Key words: **Nucleopolyhedrovirus, *Antheraea pernyi*, genome, insertion sequence, transposase**

Introduction

Lepidopteran nucleopolyhedroviruses (NPVs), classified in the genus Alphabaculovirus of the family *Baculoviridae* according to the recently proposed classification (Jehle *et al.,* 2006), have been used as viral insecticides against lepidopteran pests and as eukaryotic gene expression vectors. We have established a baculovirus expression vector (BEV) system using *Antheraea pernyi* NPV (AnpeNPV) for the high level expression of foreign genes in cultured NISES-AnPe-428 (AnPe) cells derived from *A. pernyi* embryos (Inoue and Hayasaka, 1995) as well as in larvae and diapausing pupae of wild silkmoths such as *A. pernyi* and *Samia cynthia pryeri* (Wang *et al.,* 2000; Kobayashi *et al.,* 2001). When comparing to other BEV systems such as *Autographa californica* NPV vector with Sf9 and High5 cells and *Bombyx mori* NPV vector with BmN4 cells and *B. mori* larvae, the AnpeNPV-infected diapausing puape of *S. c. pryeri* showed the highest protein production efficiency (Huang *et al.,* 2001; Kobayashi, 2001) In addition, it has been found that the structure of some *N*-glycans added to the recombinant glycoprotein by AnPe cells is bianntenary complex type which is not detected in Sf9 cells but typical in mammalian cells, suggesting that AnPe is well suited for producing pharmaceutical glycoproteins with mammlian-like *N*-glycans (Nagaya *et al.,* 2002, 2003).

A comparative genome map of AnpeNPV aligned with the fully sequenced *Orgyia pseudotsugata* multicapsid NPV (OpMNPV) genome (Ahrens *et al.,* 1997) was constructed and used for identifying and cloning several virus genes to improve protein production capabilities as well as expand host range by targeting or modifying them. (Huang *et al.,* 2002; Huang and Kobayashi, 2003; Maegawa *et al.,* 2003; 2004). For further improvements of the AnpeNPV vector system by investigating and manipuiating virus genes, we have been sequencing 54 *Pst*I fragments composing whole genome of the AnpeNPV clone A (Huang *et al.,* 2002). Recently, complete genome sequences of two different AnpeNPV isolates from Liaoning strain (GenBank accession numbers, NC_008035 and EF207986) were published (Nie *et al.,* 2007; Fan *et al.,* in press), enabling us to compare genome sequences among isolates, although the sequence of the latter isolate (EF207986) is not available during writing this paper.

In this paper, we describe a truncated ORF in a mutable region found in *Pst*I D fragment of clone A by comparing with the corresponding sequence of the Liaoning strain (NC_008035) as well as several closely related NPVs. In addition, two mutants of *Pst*I D fragment generated during propagation in bacterial cells are also reported.

Materials and Methods

Bacterium and DNA

Competent *E. coli* strain XL1-Blue cells (strangene) were used for the transformation and propagation of plasmid DNA. *Pst*I D fragment of the AnpeNPV clone A genome DNA cloned in pBluescriptII (Stratagene) (Huang *et al.,* 2002) were used for nucleotide sequence determination.

DNA manipulations and sequencing

All plasmind DNA recombination techniques were essentially as described by Sambrook *et al.* (1989). Restriction enzymes and other DNA modifying enzymes were purchased from Takara Bio Co Ltd. The nucleotide sequence of *Pst*I D fragment of the AnpeNPV was determind by Dragon Genomics Center of Takara Bio Co Ltd.

Sequence analysis

Homology search of nucleotide and predicted amino asid sequences of *Pst*I D fragment was carried out using the basic local alignment search tool (BLAST) program provided by the GenomeNet WWW Server (Bioinformatics Center, Kyoto University). The GENETYX program (Genetyx) was used for both sequence data analysis and alignment of sequences from AnpeNPV clone A and Liaoning strain as well as two mutants of *Pst*I

D fragment.

Nucleotide sequence accession number

The nucleotide sequence of the total 6318 bp *Pst*I D fragment of AnpeNPV clone A containing 11 ORFs (2 partial and 9 complete ORFs) has been submitted to the DDBJ/EMBL/Genbank databases under accession number AB353738.

Results and Discussion

Comparative sequence analysis of Pst*I D fragment of AnpeNPV clone A*

Nucleotide sequence of 6318 bp *Pst*I D fragment of AnpeNPV clone A was determined. As shown in Fig. 1, eleven open reading frames (ORFs) (9 complete and 2 partial ORFs) corresponding to Anpe111 (*pif-1*), 110, 109, 108 (*bro-a*), 107 (*pif-3*), 106, 105, 104 (*Baculo_8kD*), 103, 102 (*ChtBD-2*) and 101 of the AnpeNPV Liaoning strain (GenBank accession number NC_008035) were found in the fragment. Except one ORF corresponding to Anpe105, positions, directions and lengths of the other 10 ORFs of clone A were completely or almost identical to the corresponding ORFs of the Liaoning strain, respectively (Table 1). Pair wise comparison with the Liaoning strain genome sequence revealed 8 base substitutions in the *Pst*I D fragment of clone A, resulting in one amino acid replacement which cause slight decreases in identity (between 99.5 and 99.8%) within each of putative proteins encoded dy 4 ORFs corresponding to Anpe101, 106, 107 and 111. In addition, an insertion (4bp) and a deletion (10 bp) were identified at 6 nucleotides (nt) upstream and 50 nt downstream from A of ATG translational start codon of the ORF corresponding to Anpe105, respectively (Fig. 2). Although the insertion does not alter the TAAG late promoter element, the deletion causes a frameshift and the earlier termination of the coding sequence, resulting in the truncation of Anpe 105 product (71 aa) to a much smaller polypeptide (22 aa).

Computer-assisted homology seaches of amino acid sequences revealed the presence of Anpe105 homologs in genomes of several closely related NPVs such as *Anticarsia gemmatalis* NPV(AgNPV) ORF 111 (Oliveira *et al.*, 2006), *Choristoneura fumiferana* defective NPV (CfDEFNPV) ORF 107 (Lauzon *et al.*,

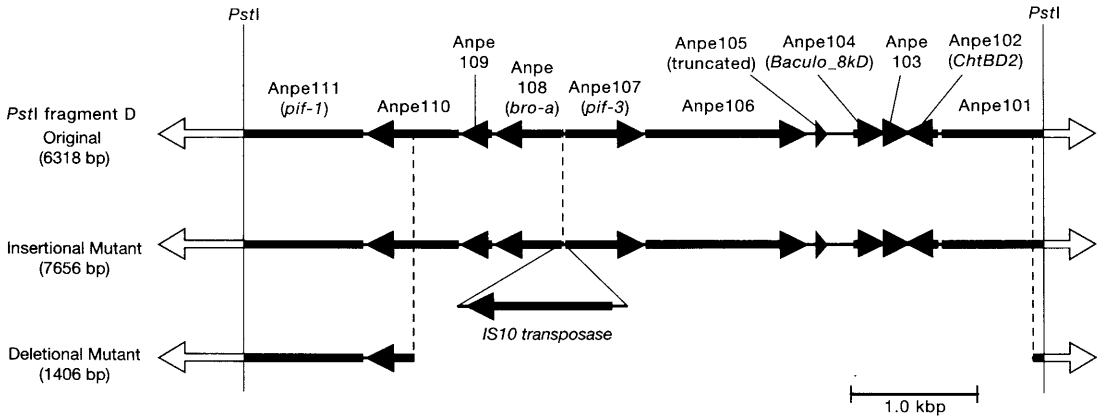

Fig. 1. Gene arrangement map of the AnpeNPV clone A *Pst*I D fragment (top) and its insertional and deletional mutants (middle and bottom). The positions and directions of 11 ORFs corresponding to the Liaoning strain Anpe111(*pif-1*), 110, 109, 108(*bro-a*), 107(*pif-3*), 106, 105, 104(*Baculo_8kD*), 103, 102(*ChtBD2*) and 101 are shown by black arrows. For the insertional mutant(middle), the inserted position and length of IS10 are indicated by a bar with a black arrow showing the position and direction of the *transposase* gene underneath the fragment, while deleted region is erased for the deletion mutant(bottom). *Pst*I sites at the both ends of each fragment are indicated. A scale bar represents 1.0 kbp.

Table 1 AnpeNPV *Pst*I D fragment ORFs

ORF	Name	clone A			L strain			Number of base substitution	Number of deletion/ insertion	Amino acid identity % (Range)
		Position	Amino acids	Intergenic distance	Position	Amino acids	Intergenic distance			
Anpe111*	*pif-1*	...1 < 954	318		90968 > 91921...	318		1	0	99.7 (318)
				41			41	0	0	
Anpe110		996 < 1706	236		90216 > 90926	236		0	0	100 (236)
				24			24	0	0	
Anpe109		1731 < 1976	81		89946 > 90191	81		0	0	100 (81)
				29			29	0	0	
Anpe108	*bro-a*	2006 < 2539	147		89383 > 89916	147		0	0	100 (147)
				27			27	0	0	
Anpe107	*pif-3*	2567 > 3184	205		88738 < 89355	205		1	0	99.5 (205)
				8			8	1	0	
Anpe106		3193 > 4443	416		87479 < 88729	416		3	0	99.8 (416)
				105			101	0	1 (+4 bp)**	
Anpe105		4549 > 4617	22		87162 < 87377	71		0	1 (-10 bp)**	72.7 (22)
				205			68	0	0	
Anpe104	*Baculo_8kD*	4823 > 5071	82		86845 < 87093	82		1	0	100 (205)
				-4			-4	0	0	
Anpe103		5068 > 5238	56		86678 < 86848	56		0	0	100 (56)
				-9			-9	0	0	
Anpe102	*ChtBD2*	5230 < 5508	92		86408 > 86686	92		0	0	100 (92)
				21			21	0	0	
Anpe101*		5530 > 6318...	262		...85598 < 86386	262		1	0	99.6 (262)

AnpeNPV ORF is indicated in 'ORF' column according to the Liaoning (L) strain (NC_008035). The right and left boundaries are given by nucleotide number in the 'postion' column and the direction of transcription is given by the <(-strand) and> (+strand) symbols. The intergenic distance between the ORFs is given in bp, negative number indicates overlap. *; partial ORF. **; length (bp) of deletion (-) or insertion (+) in clone A.

2005), *Epiphyas postvittana* NPV (EppoNPV) ORF 98 (Hyink *et al.*, 2002), *Hyphantria cunea* NPV (HycuNPV) ORF 43 (Ikeda *et al.*, 2006), *Choristoneura fumiferana* NPV (CfNPV) ORF 103 (de Jong *et al.*, 2005) (Fig. 3). These homologs are longer (121-146 aa) than Anpe105 and have a domain with highly identical amino acid sequences including 4 conserved cysteine residues from amino acid position 70 to 100. Anpe105 product almost completely lacks the conserved domain. The results of comparative sequence analysis suggested that these Anpe105 homologs are derived from an ancestor gene acquired in the phylogenetic subgroup of group I NPVs (Herniou *et al.*, 2003) that includes AnpeNPV, AgNPV, CfNPV, EppoNPV, HycuNPV, CfDEFNPV and OpMNPV. Although any functions have not been elucidated their products yet, Anpe105 products lacking the conserved domain are probably nonessential or functionless for the replication of AnpeNPV, as indicated by the facts that AnpeNPV clone A with a truncated Anpe105 can normally propagate and express heterologous genes in *A. pernyi* larvae, pupae and cell culture (Wang *et al.*, 2000, Huang *et al.*, 2001).

OpMNPV, a member of the subgroup of group I NPVs, lacks Anpe105 homolog in the corresponding region of the genome, but have a different ORF (OpORF113) instead (Ahrens *et al.*, 1997). Curiously, each member of the subgroup except AnpeNPV also has a homolog of OpORF113 in the corresponding region between Anpe104 and 105 homologs. Probably OpMNPV lost Anpe105 homolog and AnpeNPV lost OpORF 113 homolog respectively during the diversifiation from the subgroup ancestor. In addition, AnpeNPV may be just losing Anpe105. Sequence heterogeneity in the Anpe105 region among AnpeNPV isolates seemed to support the possibility.

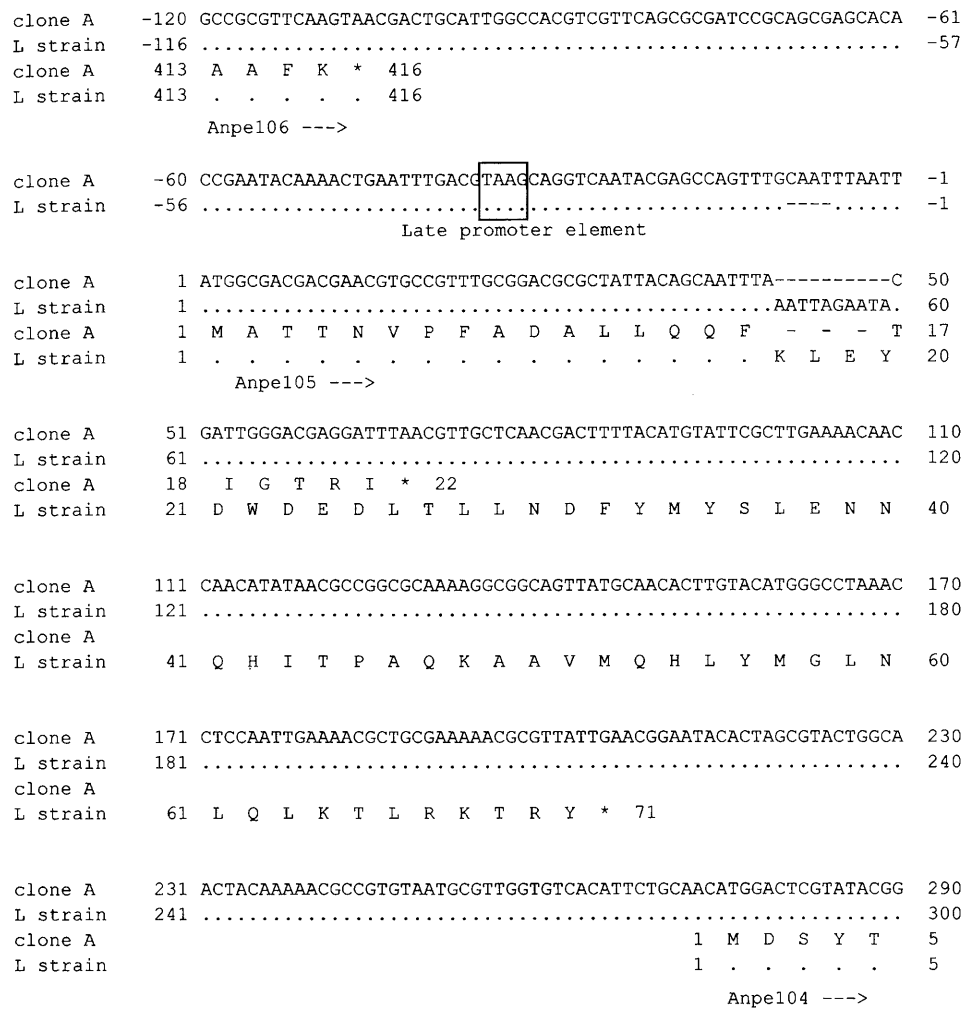

Fig. 2. Comparison of nucleotide sequences of Anpe105 and surrounding region between the AnpeNPV clone A(above) and Liaoning (L) strain (NC_008035) (below). The first nucleotide of the ATG translation initiation codon has been designated number +1. The deduced amino acid sequences of Anpe105 as well as of Anpe104 and 106 are also shown beneath the nucleotide sequences with amino acid number from the first methionine. The late promoter elements(TAAG) at upstream of Anpe105 are boxed and arrows indicate translational directions of respective ORFs. Dots indicate identical nucleotides and amino acids and dashes indicate gaps.

Two mutants of Pst*I D fragment generated in bacterial cells*

For the sequence determination of *Pst*I D fragment cloned into pBluescriptII, colonies derived from competent cells of *Escherichia coli* transformed with the plasmind DNA stock were selected and propagated for the large-scale preparation of the plasmind. During the colony selection, two colonies harboring mutant plasmids differing in length were obtained. Comparative sequence analysis has revealed that a longer mutant has a 1329-bp of bacterial insertion sequence IS10 and a direct duplication of a 9-bp target sequence (5'-TGCTGCGCC-3') at intergenic region between *bro-a* and *pif-3* (Fig. 1). The IS10 sequence consists of a 22-bp inverted repeat at the ends and an ORF encoding IS10 transposase (402

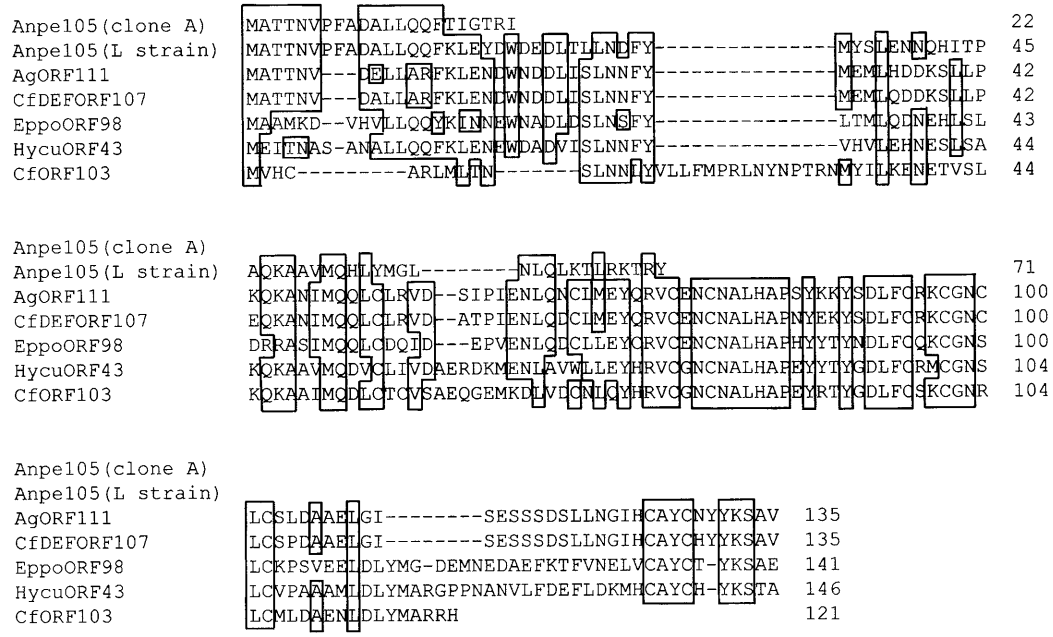

Fig. 3. Amino acid sequence alignment of Anpe105 of two AnpeNPVs (clone A and Liaoning (L) strain) with its homologs in other group I NPVs, AgNPV ORF 111(Oliveira *et al.*, 2006), CfDEFNPV ORF107 (Lauzon *et al.*, 2005), EppoNPV ORF 98 (Hyink *et al.*, 2002), HycuNPV ORF 43 (Ikeda *et al.*, 2006) and CfNPV ORF 103 (de Jong *et al.*, 2005). Amino acid residues conserved in more than 4 virus sequences are boxed. Dashes indicate gaps. Amino acid number is indicated on the right.

aa). A shorter mutant is resulted from a 4912-bp deletion containing 11 ORFs corresponding to Anpe101 to 110 (Fig. 1). PCR analyses of AnpeNPV genome DNA using primers specific to *Pst*I D fragment and IS10 sequences clearly indicated that these two mutant fragments were not contained in the virus DNA preparation but created in the bacterial cells (data not shown).

The contamination of prokaryotic insertion sequence (IS) has been reported in sequences of eukaryotes and eukaryotic viruses registered in databases (Binns *et al.* 1986, Binns 1993, Astua-Monge *et al.* 2002) These reports also suggested that IS elements were frequently associated with deletion and other genetic rearrangements. Therefore, it is quite likely that the IS10 element transposed from *E. coli* chromosome to the plasmid carriying *Pst*I D fragment shortly after transformation and mediated delection of the fragment in some plasmids, although deletions could occur independently of IS10. Such alterations in cloned sequences are very troublesome for the sequence determination. In fact, we performed confirmation sequencing of *Pst*I D fragment using different clones.

Acknowledgments

This work was partly supported by a grant for Insect Factory Research Project from National Institute of Agrobiogical Sciences, Japan.

References

Ahrens, C. H., R. L. Russell, C. J. Funk, J. T. Evans, S. H. Harwood and G. F. Rohrmann, 1997. The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. Virology, **229**: 381-99.

Astua-Monge, G., A. Lyznik, V. Jones, S. A. Mackenzie and C. E. Vallejos, 2002. Evidence for a prokaryotic insertion-

sequence contamination in eukaryotic sequences registered in different databases. Theor. Appl. Genet., **104**: 48-53.

Binns, M., 1993. Contamination of DNA database sequence entries with *Escherichia coli* insertion sequensces. Nucleic Acids Res., **21**: 779.

Binns, M. M., M. G. Boursnell, F. M. Tomley and D. K. Brown, 1986. Comparison of the spike precursor sequences of coronavirus IBV strains M41 and 6/82 with that of IBV Beaudette. J. Gen. Virol., **67**: 2825-2831.

de Jong, J. G., H. A. M. Lauzon, C. Dominy, A Popoumienko, E. B. Carstens, B. M. Arif and P. J. Krell, 2005. Analysis of the *Choristoneura fumiferana* nucleopolyhedrovirus genome. J. Gen. Virol., **86**: 929-943.

Fan, Q., S. Li, L. Wang, B. Zhang, B. Ye, Z. Zhao and L. Cui, 2007. The genome sequence of the multinucleocapsid nucleopolyhedrovirus of the Chinese oak silkworm *Antheraea pernyi*. Virology, **360**: 235-246.

Herniou, E. A., J. A. Olszewski, J. S. Cory and D. R. O'Reilly, 2003. The genome sequence and evolution of baculoviruses. Annu. Rev. Entomol., **48**: 211-234.

Huang, Y. J., X. Y. Wang, S. Miyajima, T. Yoshimura, and J. Kobayashi, 2001. Efficiency of *Antheraea pernyi* nucleopolyhedrovirus-mediated protein production in both an established cell line and diapausing pupae of *A. pernyi*. Int. J. Silkmoth & Silk, **6**: 59-71.

Huang, Y. J. and J. Kobayashi, 2003. *Antheraea pernyi* nucleopolyhedrovirus mutants lacking functional cathepsin and/or chitinase genes suppress proteolysis and liquefaction of virus-infected diapausing *A. pernyi* pupae. Int. J. Wild Silkmoth & Silk, **8**: 43-50.

Huang, Y. J., J. Kobayashi, and T. Yoshimura, 2002. Genome mapping and gene analysis of *Antheraea pernyi* nucleopolyhedrovirus for improvement of baculovirus expession vector system. J. Biosci. Bioeng., **93**: 183-191.

Ikeda, M., M. Shikata, N. Shirata, S. Chaeychomsri and M. Kobayashi, 2006. Gene organization and complete sequence of the *Hyphantria cunea* nucleopolyhedrovirus genome. J. Gen. Virol., **87**: 2549-2562.

Inoue, H. and S. Hayasaka, 1995. A new cell line separated from the contractile muscle cell line of Chinese oak silkworm, *Antheraea pernyi*. J. Seric. Sci. Jpn., **64**: 79-81.

Jehle, J. A., G. W. Blissard, B. C. Bonning, J. S. Cory, E. A. Herniou, G. F. Rohrmann, D. A. Theilmann, S. M. Thiem and J. M. Vlak, 2006. On the classification and nomenclature of baculoviruses: a proposal for revision. Arch. Virol., **151**: 1257-1266.

Kobayashi, J., 2001. Developing insect gene expression system in the genome era. Proc. Joint Int. Symp. "Prospect for the Development of Insect Factories", 49-54.

Kobayashi J., R. Ando, X. Y. Wang, Y. J. Huang and S. Miyajima, 2001. Nucleotide sequence analysis of the polyhedrin gene of *Antheraea penyi* nucleopolyhedrovirus and construction of a transfer vector plasmid. Int. J.Wild Silkmoth & Silk, **6**: 53-58.

Lauzon, H. A. M., P. B. Jamieson. P. J. Krell and B. M. Arif, 2005. Gene organization and sequencing of the *Christoneura fumiferana* defective nucleopolyhedrovirus genome. J.Gen. Virol., **86**: 945-961.

Maegawa, K., J. Kobayashi and T. Yoshimura, 2003. *Antheraea pernyi* nucleopolyhedrovirus p143 gene: nucleotide sequence and ineffectiveness in host range expansion of *Autugrapha california* nucleopolyhedrovirus. Int. J. Wild Silkmoth & Silk, **8**: 29-41.

Maegawa, K., X. Y. Wang, J. Kobayashi and T. Yoshimura, 2004. Nucleotide sequence analysis of *p10* gene of *Antheraea pernyi* nucleopolyhedrovirus and construction of two transfer vector plasmids. Int. J. Wild Silkmoth & Silk, **9**: 53-60.

Nagaya, M., J. Kobayashi, N. Takahashi, K. Kato and T. Yoshimura, 2003. Two-dimensional high performance liquid chromatography mapping of sugar chains demonstorated the biantennary, complex *N*-glycan addition to the recombinant glycoprotein produced by baculovirus-infected *Antheraea pernyi* insect cells. J. Insect Biotechnol. Sericol., **72**: 79-86.

Nagaya, M., J. Kobayashi and T. Yoshimura,

2002. Evaluation of N-glycosylation property of cultured *Antheraea pernyi* cells in comparison with four other lepidopteran cell lines used in beculovirus expression vector systems. Int. J. Wild Silkmoth & Silk, **7**: 59-68.

Nie, Z.-M., Z.-F. Zhang, D. Wang, P.-A. He, C.-Y. Jiang, L. Song, F. Chen, J. Yang, L.-L. Yu, J. Chen, Z.-B. Lv, J.-J. Lu, X.-F. Wu and Y.-Z. Zhan, 2007. Complete sequence and organization of *Antheraea pernyi* nucleopolyhedrovirus, a *dr*-rich baculovirus. BMC Genomics, **8**: 248.

Oliveira, J. V., J. L. Wolff, A. Garcia-Maruniak, B. M. Ribeiro, M. E. de Castro, M. L. de Souza, F. Moscardi, J. E. Maruniak and P. M. Zanotto, 2006. Genome of the most widely used viral biopesticide: *Anticarsia gemmatalis* multiple nucleopolyhedrovirus. J. Gen. Virol., **87**: 3233-3250.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, New York, NY.

Wang, X. Y., J. Kobayashi, S. Miyajima, H. J. Xie, S. L. Ghi, W. Y. Zheng and R. Q. Ji, 2000. Cloning and property analysis of *Antheraea pernyi* nucleopolyhedrovirus (AnpeNPV). Int. J. Wild Silkmoth & Silk, **5**: 51-56.

摘　要

サクサン核多角体病ウイルス（AnpeNPV）プラーク純化クローンA株の*Pst*I D 断片（6318bp）の塩基配列が決定され，全ゲノム配列が最近公開された遼寧株（GenBank 登録番号NC_008035）と比較された。断片中に同定された11個のORFは，遼寧株のAnpe101から111に対応しており，Anpe105を除き完全あるいはほとんど同一であった。Anpe105に対するクローンA株のORFは，翻訳配列内に生じた10塩基分の欠失によるフレームシフトと早期の翻訳終止により短縮していた。さらなる比較ゲノム解析により，いくつかのグループIに属するNPVがAnpe105 のホモログを有することが明らかとなった。これら全てのホモログの産物は，Anpe105産物よりも非常に長く，Anpe105では欠落している保存領域を共有していた。以上の結果は，Anpe105がAnpeNPVの増殖サイクルにおいて非必須もしくは不活性であり，それゆえに，Anpe105領域のゲノム配列が不安定で，挿入や欠失を受けやすいことを示唆した。また，クローニングした*Pst*I D 断片を含むプラスミドを大腸菌で増幅させる際に，2種類の変異が得られた。配列分析により，一方の変異では，Anpe107と108の間に相当する領域にIS10トランスポゾン（1329bp）が挿入されており，他方の変異では，ApORF101から110に相当する領域（4912bp）が欠損していることがわかった。これらの変異について考察した。