Virologica Sinica (2018) 33:359–368
https://doi.org/10.1007/s12250-018-0047-9

www.virosin.org
www.springer.com/12250

RESEARCH ARTICLE

Genome Characteristics of the Cyclophragma Undans Nucleopolyhedrovirus: A Distinct Species in Group I of *Alphabaculovirus*

Zheng Zhu[1] · Jun Wang[1] · Qianran Wang[1] · Feifei Yin[1] · Xiaoping Liu[1] · Dianhai Hou[1] · Lei Zhang[1] · Haizhou Liu[1] · Jiang Li[1] · Basil M. Arif[2] · Hualin Wang[1] · Fei Deng[1] · Zhihong Hu[1] · Manli Wang[1]

Received: 31 May 2018 / Accepted: 16 July 2018 / Published online: 28 August 2018

Abstract

The Cyclophragma undans nucleopolyhedrovirus (CyunNPV), a potential pest control agent, was isolated from *Cyclophragma undans* (Lepidoptera: Lasiocampidae), an important forest pest. In the present study, we performed detailed genome analysis of CyunNPV and compared its genome to those of other Group I alphabaculoviruses. Sequencing of the CyunNPV genome using the Roche 454 sequencing system generated 142,900 bp with a G + C content of 45%. Genome analysis predicted a total of 147 hypothetical open reading frames comprising 38 baculoviral core genes, 24 lepidopteran baculovirus conserved genes, nine Group I *Alphabaculovirus* conserved genes, 71 common genes, and five genes that are unique to CyunNPV. In addition, the genome contains 13 homologous repeated sequences (*hrs*). Phylogenetic analysis groups CyunNPV under a distinct branch within clade "a" of Group I in the genus *Alphabaculovirus*. Unlike other members of Group I, CyunNPV harbors only nine of the 11 genes previously determined to be specific to Group I viruses. Furthermore, the CyunNPV lacks the tyrosine phosphatase gene and the *ac30* gene. The CyunNPV F-like protein contains two insertions of continuous polar amino acids, one at the conventional fusion peptide and a second insertion at the pre-transmembrane domain. The insertions are likely to affect the fusion function and suggest an evolutionary process that led to inactivation of the F-like protein. The above findings imply that CyunNPV is a distinct species under Group I *Alphabaculovirus*.

Keywords Cyclophragma undans nucleopolyhedrovirus (CyunNPV) · Baculovirus · Group I *Alphabaculovirus* · F protein

Introduction

Cyclophragma undans (Lepidoptera: Lasiocampidae) is an important forest pest that has caused serious damage to Masson pines, oaks, and Tung oil trees that are commonly found in China (Xiao and Shu 1984). Cyclophragma undans nucleopolyhedrovirus (CyunNPV) was first isolated from field-diseased *Cyclophragma undans* larvae and was characterized as a multiple nucleocapsid nucleopolyhedrovirus (Huang *et al*. 1983).

Baculoviridae is a family of insect viruses with genome sizes ranging from 81,755 bp (Neodiprion lecontei nucleopolyhedrovirus, NeleNPV) to 178,733 bp (Xestia c-nigrum granulovirus, XecnGV) (Hayakawa *et al.* 1999; Lauzon *et al.* 2004). The *Baculoviridae* family comprises four genera, namely, *Alphabaculovirus* (NPVs of lepidopteran insects), *Betabaculovirus* [granuloviruses (GVs) of Lepidoptera], *Gammabaculovirus* (NPVs of Hymenoptera), and *Deltabaculovirus* (NPVs of Diptera) (Carstens and Ball 2009; Jehle *et al.* 2006). *Alphabaculovirus* is further divided into Groups I and II NPVs based on phylogeny (Bulach *et al.* 1999; Jehle *et al.* 2006; Zanotto *et al.*

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12250-018-0047-9) contains supplementary material, which is available to authorized users.

✉ Manli Wang
wangml@wh.iov.cn

[1] State Key Laboratory of Virology and China Center for Virus Culture Collection, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, China

[2] Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste Marie, ON P6A 2E5, Canada

1993). Viruses under Group I are considered relatively newly diverged in *Baculoviridae* (Herniou *et al.* 2003; Jiang *et al.* 2009). Unlike other lepidopteran baculoviruses that contain the fusion (F) protein as the primary membrane protein, Group I viruses utilize GP64 as the membrane fusion protein. The acquisition of GP64 has been suggested to be the major evolutionary event that facilitated the radiation of Group I alphabaculoviruses (Jiang *et al.* 2009; Pearson *et al.* 2000).

To date, the complete genomes of over 80 baculoviruses have been sequenced and deposited in the NCBI database. Genome analysis identified 38 conserved genes among all baculoviruses that were designated as the core baculovirus genes (Garavaglia *et al.* 2012). Furthermore, 24 genes are shared among all sequenced lepidopteran baculoviruses (Garavaglia *et al.* 2012; van Oers and Vlak 2007), and 11 genes are unique to Group I viruses (Herniou *et al.* 2003; Jiang *et al.* 2009).

In the present study, sequencing and phylogenetic analysis of the CyunNPV genome shows that it has evolved early within Group I. Unlike the majority of Group I viruses, CyunNPV lacks two of the 11 genes considered unique to Group I. In addition, the F-like protein of CyunNPV is diverged relative to the other F and F-like proteins, and this difference can provide insights underlying the divergence of the F-like protein from the F protein. Our findings provide evidence that CyunNPV is a distinct species in Group I.

Materials and Methods

Viral DNA Extraction

The CyunNPV occlusion bodies (OBs) were isolated from infected *Cyclophragma undans* larvae and purified by differential and rate zonal centrifugation (Deng *et al.* 2007). The OBs of CyunNPV were preserved in 50% glycerol and stored in − 40 °C in the Culture Collection Center of the Wuhan Institute of Virology, Chinese Academy of Sciences (preserved ID: IVCAS 1.0135). Viral DNA was extracted directly from the purified OBs by dissolution in 0.17 mol/L NaCl, 0.01 mol/L EDTA, and 0.1 mol/L Na_2CO_3. Afterwards, extracted DNA was treated with proteinase K, SDS, and phenol–chloroform and subsequently subjected to dialysis (Hu *et al.* 1998b).

Sequencing and Assembly

CyunNPV DNA (500 ng) was used for Roche 454 GS-FLX-Titanium library preparation (fragment size: > 1.5 kb) and shotgun sequencing using the Roche 454 FLX sequencing system. Reads were filtered using Newbler to obtain high-quality reads for assembly. Genome contigs were assembled using the Roche GS De Novo assembler software Newbler 2.6 with default parameters. Gene collinearity was used to predict the relationships among the contigs. Gap filling was performed via PCR, and sequences verified by Sanger sequencing. The complete genome sequence and annotation information were deposited in GenBank (accession number: KT957089).

Bioinformatics Analysis

Predicted ORFs were identified using the Softberry FGENESV program (http://www.softberry.com/berry.phtml) (Solovyev and Salamov 1999). ORFs containing less than 50 amino acids were filtered out. Gene-parity plot analysis was conducted as previously described (Hu *et al.* 1998a; Zhu *et al.* 2014). Gene annotation and comparisons and detection of duplicate sequences were performed using NCBI's BLAST algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Restriction sites were predicted using DNAstar MapDraw software. Genome maps were generated using genomeVX (Conant and Wolfe 2008). MEGA 6 software was used to perform multiple alignment. The phylogenetic tree was constructed based on 38 core genes or the *gp64* gene based on maximum likelihood method using the JTT matrix-based model (Jones *et al.* 1992; Tamura *et al.* 2013). Transmembrane domains were predicted using the TMHMM online server (http://www.cbs.dtu.dk/services/TMHMM/).

Results

Sequencing and Genome Characteristics

Shotgun sequencing of the complete CyunNPV genome was conducted using a Roche 454 system. Scaffolds were assembled using Roche GS De Novo assembler software using high-quality reads. The genome was sequenced at 343× coverage with 138,528 reads and an average read - length of 331 bp. Gaps in the genome scaffolds were filled via PCR and Sanger sequencing. The genome sequence was 142,900 bp in length and had a G + C content of 45%. Genome analysis identified 147 methionine-initiated ORFs containing at least 150 nt with minimal overlaps (Supplementary Table S1). The *polyhedrin* gene is set as the first ORF (ORF 1) in the circular genome (Fig. 1). The 38 core genes (red), 24 conserved genes in lepidopteran baculoviruses (blue), and the nine Group I unique genes (green) are illustrated on the circular map (Fig. 1). Two of the previously identified Group I unique genes, namely, protein tyrosine phosphatase gene (*ptp*) and *ac30,* were not detected in CyunNPV. The remaining 71 common genes and five unique genes are shown in gray as open arrows (Fig. 1).

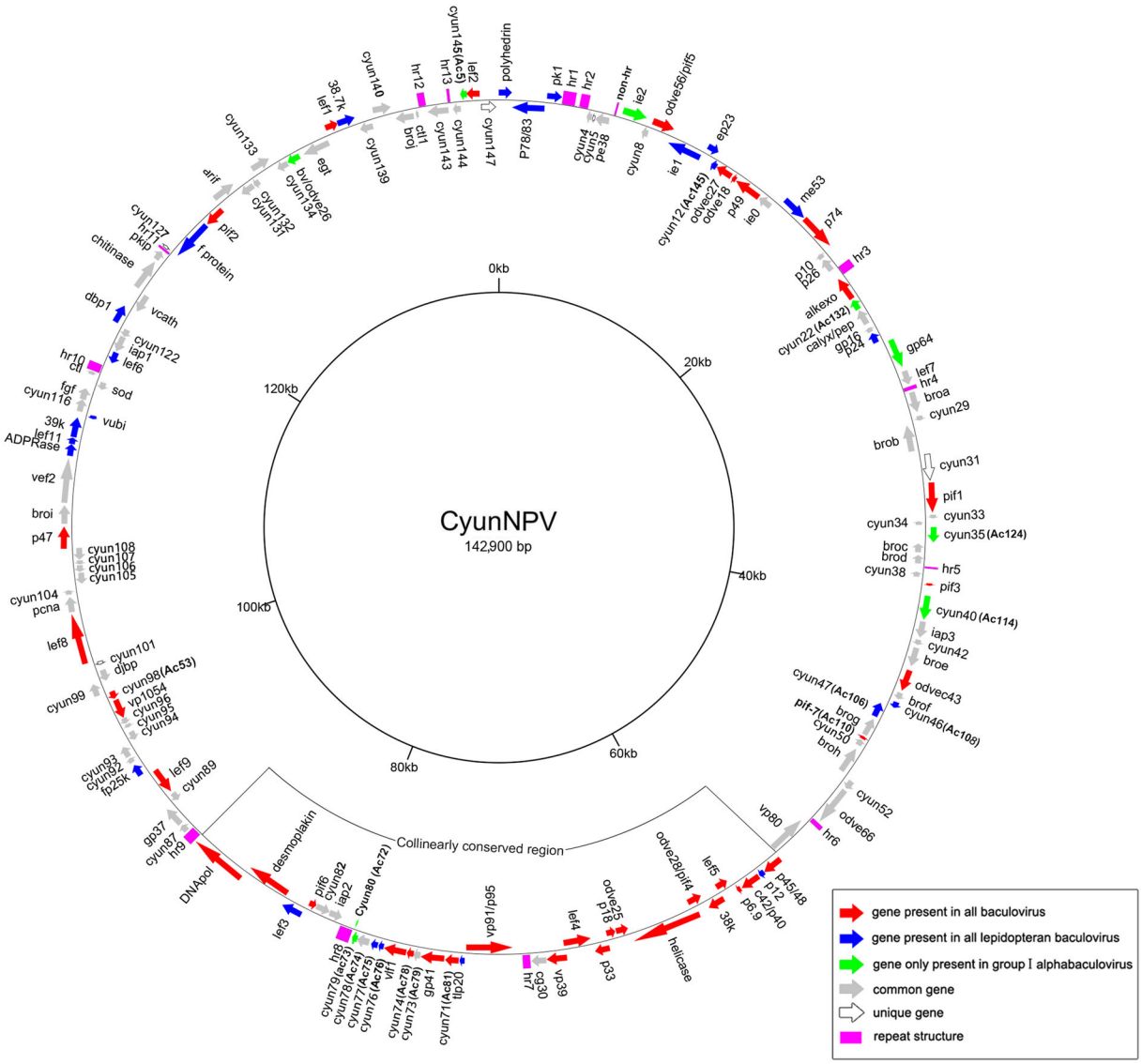

Fig. 1 Gene organization of CyunNPV. Arrows depict an indicated ORF and its transcriptional direction. The colors represent gene types: red indicates core genes, blue indicates genes conserved in lepidopteran baculoviruses, green indicates genes unique to members of Group I *Alphabaculovirus*, gray indicates common genes, and open arrows indicate genes unique to CyunNPV. Pink squares represent repeat structures. Inner circle indicates genome scale position by 20 kb. The collinearly conserved region of lepidopteran baculoviruses is also indicated.

Among the 147 ORFs, 118 genes were categorized as follows based on function: 13 replication-associated ORFs, 11 transcription-related ORFs, ten ORFs involved in oral infection, 34 ORFs related to structure, and 50 auxiliary genes (Table 1). The functions of the remaining 29 ORFs remain to be determined.

Repeat Regions

Most baculoviral genomes contain homologous repeat sequences (*hr*s), which typically consist of a few repeated sequences with imperfect palindromes that are interspersed within the genome. The *hrs* of different baculoviruses are highly variable (Ferrelli *et al.* 2012), and previous studies have suggested that *hrs* can act as origins of DNA replication and enhancers of gene expression (Guarino *et al.* 1986; Kool *et al.* 1993; Rodems and Friesen 1995; Theilmann and Stewart 1992). A total of 13 *hrs* were identified in the CyunNPV genome (Fig. 2A). A cAMP response element (CRE)-like motif and its palindrome sequences were found in the *hr*s of CyunNPV (Fig. 2A). Additionally, an AT-rich region and a GC-rich region were detected in the *hrs* of CyunNPV (Fig. 2A). Transient assays revealed that the CRE-like motif functions as a transcriptional activator in AcMNPV (Landais *et al.* 2006). Non-*hr* replication origins have been reported in some baculoviruses (Habib and

Table 1 Gene content of CyunNPV.

Gene types	Core genes	Lepidoptera conserved genes	Other genes
Replication	*alk-exo (Cyun21), dna polymerase (Cyun86), helicase(Cyun62), lef1(Cyun137), lef2(Cyun146)*	*ie1(Cyun10), me53(Cyun17), lef3(Cyun84), lef11(Cyun113), dbp-1(Cyun123)*	*lef7 (Cyun27), ac79 (cyun73), lef12 (Cyun108)*
Transcription	*lef4(Cyun66), lef5(Cyun59), lef8(Cyun102), lef9(Cyun90), p47(Cyun109), vlf1(Cyun75)*	*pk1(Cyun3),39 k(Cyun114), lef6(Cyun120)*	*i.e.-0(Cyun16), i.e.-2(Cyun7)*
Structure	*odv-ec27(Cyun13), odv-e18(Cyun14), p49(Cyun15), odv-ec43(Cyun44), p45/p48(Cyun55), c42/p40(Cyun57), p6.9(Cyun58),38 k(Cyun60), odv-e25(Cyun63), p18(Cyun64), p33(Cyun65), vp39(Cyun67), ac81(Cyun71), gp41(Cyun72), ac78(Cyun74),desmoplakin(Cyun85), vp1054(Cyun97), ac53(Cyun98)*	*polyhedrin(Cyun1), p78/83 (Cyun2), p24(Cyun25), p12(Cyun56), tlp-20(Cyun70), fp25 k(Cyun91), f protein(Cyun128)*	*p10(Cyun19), calyx/pep(Cyun23), gp16 (Cyun24), gp64 (Cyun26), odv-e66 (Cyun53), vp80(Cyun54), cg30 (Cyun68), pkip(Cyun126), odv-e26 (Cyun135)*
Oral infection	*pif5(Cyun9), p74(Cyun18), pif1(Cyun32), pif3(Cyun39), pif7(Cyun49), pif4(Cyun61), vp91/p95 (Cyun69), pif6(Cyun83), pif2(Cyun129)*	*ac108(Cyun46)*	
Auxiliary		*ADPRase(Cyun112), v-ubi(Cyun115), 38.7 k(Cyun138)*	*pe38(Cyun6), ac150(cyun8), p26(Cyun20),ac132(Cyun22), bro-a (Cyun28), bro-b (Cyun30), ac124 (Cyun34), bro-c (Cyun36), bro-d (Cyun37), ac117 (Cyun38), ac114 (Cyun40), iap-3(Cyun41), bro-e (Cyun43), bro-f (Cyun45), bro-g(Cyun48), bro-h(Cyun51), ac74 (Cyun78),ac73 (Cyun79),ac72 (Cyun80), iap-2(Cyun81), ac69(Cyun82), gp37(Cyun88), ac63 (Cyun89),ac60 (Cyun92),ac59 (Cyun93), ac52 (Cyun99), djbp (Cyun100), pcna (Cyun103), ac47 (Cyun104),ac45 (Cyun105),ac43 (Cyun107), bro-i(Cyun110), v- vef-2(Cyun111), ac34(Cyun116), fgf (Cyun117), sod (Cyun118), ctl (Cyun119, Cyun142), iap-1(Cyun121), ac26 (Cyun122), cath (Cyun124), chitinase(Cyun125), arif (Cyun130),ac17 (Cyun134), egt(Cyun136), bro-j(Cyun141),ac5 (Cyun145)*
Unknown		*ac75(Cyun77), ac76(Cyun76), ac106(Cyun47),ac145(Cyun12), ac146(Cyun11)*	*cyun4, cyun5, cyun29, cyun31, ac120 (Cyun33), ac122 (Cyun34), ac111 (Cyun42), cyun50, cyun52, cyun87, ac57 (Cyun94),ac56 (Cyun95),ac55(Cyun96), cyun101, ac44(Cyun106), cyun127, cyun131, ac19(Cyun132),ac18(Cyun133), ac12(Cyun139), ac11(Cyun140), cyun143,ac4 (Cyun144),cyun147*

Hasnain 2000; Kool *et al.* 1993; Pearson *et al.* 1993; Wu and Carstens 1996). The sequence features of non-*hr* are similar to those of *hr*s but appear only once within a genome. A direct repeat structure with two imperfect palindromes has been identified in the CyunNPV genome and is likely to act as a non-*hr* origin of DNA replication (Fig. 2B).

Phylogenetic Analysis Identifies CyunNPV as a Distinct Species in *Alphabaculovirus*

CyunNPV contains the *gp64* gene, indicating that it is a member of the Group I alphabaculoviruses. A phylogenetic tree was constructed using the combined 38 core genes

Fig. 2 Repeat sequences of CyunNPV. **A** Sequence comparisons of *hrs*. The four bases, namely, adenine (A), cytosine (C), guanine (G), and thymine (T), are marked as green, blue, purple, and red, respectively. The CRE-like motif and its palindromes are indicated below the alignment. **B** Sequence comparison of direct repeats. Black background shows 100% identity among the compared regions, while dark gray and light gray colors indicate 75% and 50% identity, respectively. The palindromes are indicated above the alignment. The arrows show the direction of each palindrome.

from CyunNPV, SujuNPV (Liu *et al.* 2014), ClasGV-B (Yin *et al.* 2015), and other sequenced baculoviruses in the NCBI database (Fig. 3A). CyunNPV was clustered under clade "a" of Group I and appeared to have diverged after the radiation of clade "a" (Fig. 3A).

Gene Parity Plots

Gene parity plots are useful for comparing gene organization between any two viral genomes (Hu *et al.* 1998a). CyunNPV was plotted against six representative baculoviruses (Supplementary Table S1), including the Autographa californica MNPV (AcMNPV, Group I clade "a"), Orgyia pseudotsugata MNPV (OpMNPV, Group I clade "b"), Helicoverpa armigera NPV-G4 (HearNPV-G4, Group II), Cydia pomonella granulovirus (CpGV, *Betabaculovirus*), Neodiprion lecontei nucleopolyhedrovirus (NeleNPV, *Gammabaculovirus*), and Culex nigripalpus nucleopolyhedrovirus (CuniNPV, *Deltabaculovirus*). The plots indicated that the CyunNPV gene organization was highly collinear with Group I alphabaculoviruses (Fig. 3B). Previously, we reported a collinear conserved region in lepidopteran baculoviruses comprising 20 core genes and five lepidopteran baculoviral conserved genes (Zhu *et al.* 2014). The CyunNPV genome contains the similar collinearly conserved region (red line, Fig. 3B).

CyunNPV Lacks Two Conserved Genes of Group I Viruses

Previous studies have identified 11 genes that are specific to Group I viruses and are absent from all other baculoviruses (Herniou *et al.* 2003; Jiang *et al.* 2009). These genes could have played important evolutionary roles in the divergence of Group I and Group II alphabaculoviruses (Herniou *et al.* 2003; Jiang *et al.* 2009).

Although CyunNPV belongs to Group I and contains nine of the 11 genes unique to Group I, CyunNPV lacks the protein tyrosine phosphatase gene (*ptp*) and the *ac30* gene. The protein encoded by *ptp* exhibits dual enzymatic property as a protein tyrosine phosphatase (Sheng and Charbonneau 1993) and as an RNA 5′-triphosphatase (Gross and Shuman 1998; Takagi *et al.* 1998). Deletion of *ptp* from AcMNPV is not deleterious to BV synthesis. However, *ptp* deletion impairs ODV production in Sf-21 cells but not in TN-368 cells (Li and Miller 1995). PTP appears to enhance wandering behavior in infected insects and is supported by the fact that BmNPV with a *ptp* deletion is unable to induce such behavior (Kamita *et al.* 2005). *Ac30* appears to be non-essential for BmNPV because its deletion does not affect viral production or the median lethal dose (LD_{50}), but it appears to prolong the median survival time (LT_{50}) of its host (Huang *et al.* 2008).

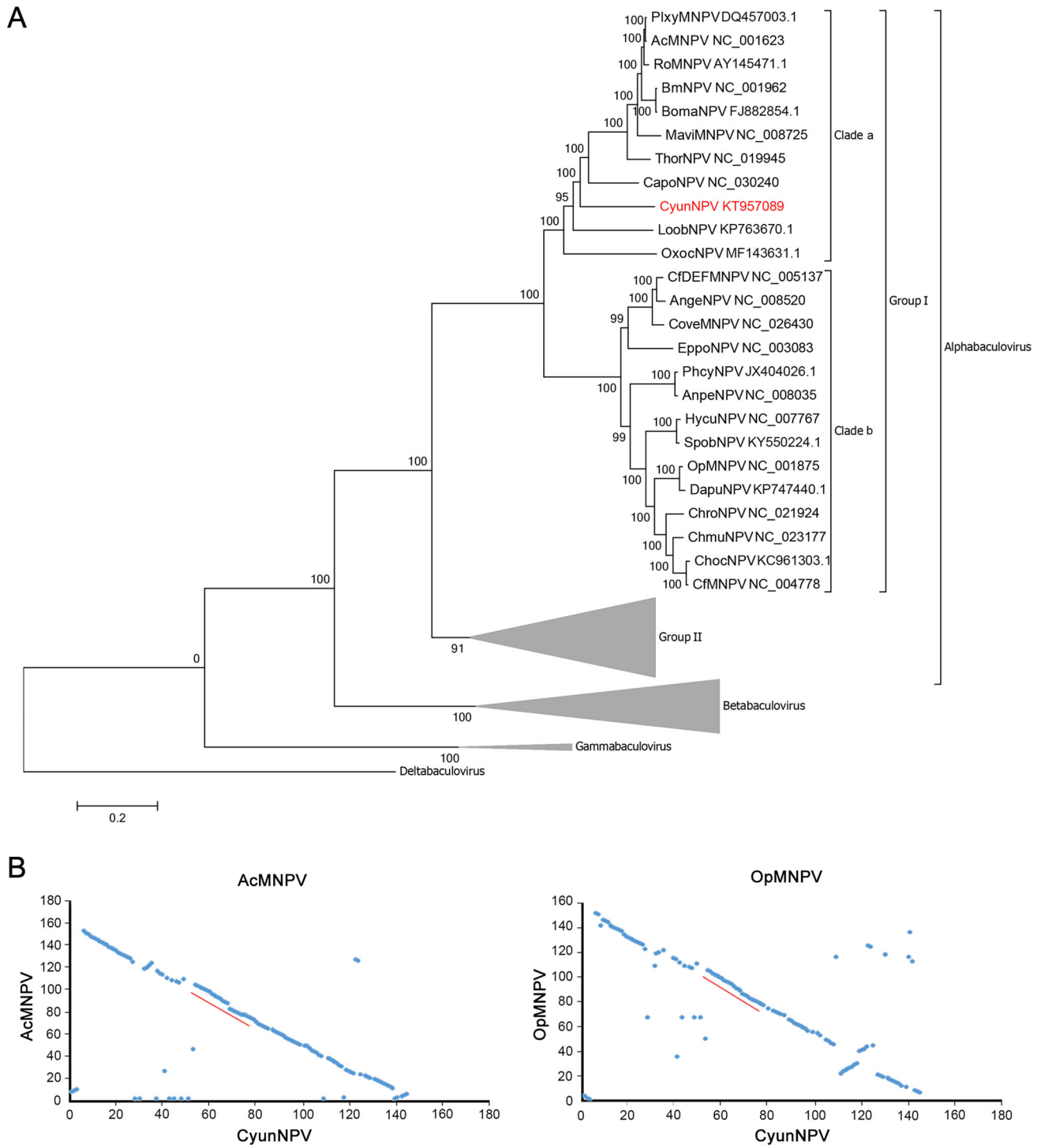

Fig. 3 Phylogenetic tree construction and gene parity plot analysis. **A** Phylogenic tree. The unrooted tree was constructed based on the 38 core proteins from all sequenced baculoviruses in the NCBI database using the maximum likelihood method. Only Group I alphabaculoviruses are shown in the phylogenetic tree. The NCBI accession numbers of the sequences of Group I alphabaculoviruses used in this study are indicated after the name of each species. The reliability of the tree was tested via the bootstrap method with a value of 500. CyunNPV is indicated in red. **B** Gene parity plot analysis. Gene parity plots of CyunNPV are compared with the Group I clade a representative species, AcMNPV, and the Group I clade b representative species, OpMNPV, based on ORF order. Red line indicates the collinearly conserved region.

There are two hypotheses that can explain the lack of the *ptp* and *ac30* genes in CyunNPV. The first hypothesis is that CyunNPV may lose these two genes during evolution, and the second hypothesis is that *ptp* and *ac30* are acquired after CyunNPV diverged from the ancestor of Group I viruses. The phylogenetic tree based on the 38 core genes grouped CyunNPV under clade "a" of Group I (Fig. 3A), thereby suggesting that the virus emerged after the split between clades "a" and "b". To further verify the above findings, a phylogenetic tree was constructed by using conserved *gp64* genes specific to Group I (Fig. 4). Similar to the above results, the generated tree grouped CyunNPV under clade "a" (Fig. 4). The ancestors of both clades independently acquire *ptp* and *ac30* after their divergence. Consequently, the *ptp* and *ac30* genes are likely to exist in the ancestor of Group I viruses, and CyunNPV has lost these two non-essential genes throughout evolution.

Sequence Analysis Provides Evidence on the Evolution of the F-Like Protein

Baculoviral F protein belongs to class I viral envelope fusion proteins, which mediate the entry of BVs into permissive cells (Pearson *et al.* 2000). Previous studies have demonstrated that the furin cleavage site is essential for protein processing and for virus/host membrane fusion (Westenberg *et al.* 2002; Ijkel *et al.* 2000). Similar to the F-like proteins found in other Group I viruses, the F-like protein of CyunNPV lacks the furin cleavage site and therefore has lost part of the fusion function (Pearson *et al.* 2000) (Fig. 5). In addition, a string of continuous polar amino acids was inserted into the fusion peptide region. Similar observations were obtained in Thysanoplusia orichalcea NPV (ThorNPV), in which five continuous glycines were inserted into the fusion peptide region

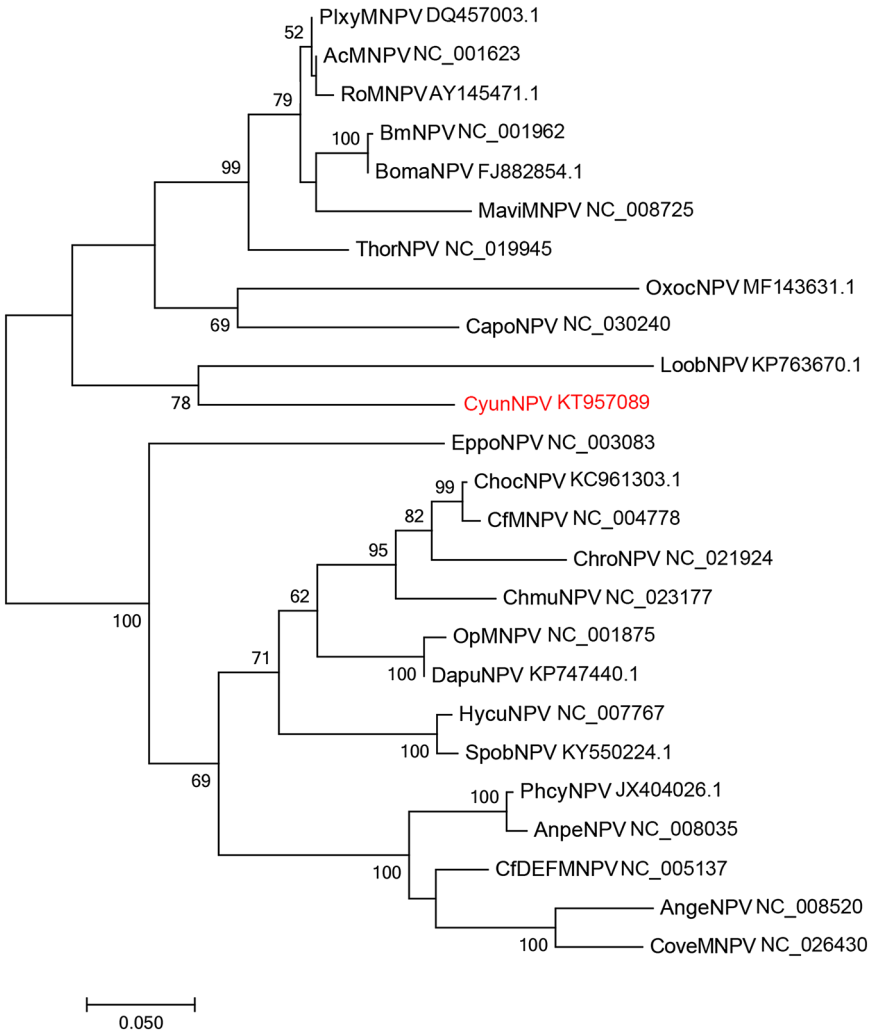

Fig. 4 Phylogenetic tree of GP64 shared in alphabaculoviruses. The unrooted tree was constructed based on the GP64 protein sequences using the maximum likelihood method. The genome accession number is indicated after each baculovirus. Bootstrap analysis was performed with a value of 500. The bootstrap values over 50% are shown in front of each node.

(Fig. 5). The fusion peptide of F proteins contains highly conserved hydrophobic amino acids that form an amphiphilic structure and play a central role in facilitating membrane fusion (Tan *et al.* 2008). These insertions are likely to inactivate the fusion peptide function of the F-like proteins of CyunNPV and ThorNPV.

Many viral fusion proteins contain a short sequence that is rich in aromatic amino acids and precedes the transmembrane domain, known as the pre-transmembrane (preTM) domain (Lorizate *et al.* 2008). The preTMs, particularly the conserved aromatic amino acids, play important roles in membrane fusion (Sainz *et al.* 2005; Salzwedel *et al.* 1999). Interestingly, the preTM regions of Group I viruses were found to contain fewer aromatic amino acids than those of other baculoviruses (Fig. 5). Strikingly, insertion of a continuous polar amino acids chain in and near the preTM domain of CyunNPV F-like protein disrupted its hydrophobicity (Fig. 5).

One of the major features differentiating Group I viruses from other baculoviruses is the presence of *gp64* and the lack of the fusion function of the F protein (*i.e.* converted to F-like). A previous study that investigated the selection pressure indicates that Group I viruses emerge from an ancient member of Group II and that the acquisition of a new fusion protein *gp64* plays an important role in the emergence of Group I viruses (Jiang *et al.* 2009). Group II F proteins are functional envelope fusion proteins that are essential for the viral fusion process (Pearson *et al.* 2000; Ijkel *et al.* 2000). Acquisition of *gp64* leads to reduced selection pressure on the F protein and the F-like proteins in Group I loses part of their function (Jiang *et al.* 2009; Pearson *et al.* 2001; Rohrmann and Karplus 2001; Wang *et al.* 2014). The gene encoding F-like protein in AcMNPV is not essential for viral infection and BV production but acts as a pathogenic factor in cellular and insect infection (Lung *et al.* 2003; Wang *et al.* 2008; Yu *et al.* 2009). It has been suggested that F-like proteins have function(s) other than membrane fusion, which could explain why they are retained in alphabaculoviruses under the evolutionary selective pressure (Wang *et al.* 2008). In this study, we identified a continuous stretch of polar amino acids inserted into the predicted fusion peptide and in the preTM region of the F-like protein. The fusion peptide and preTM regions are essential for the viral fusion function. The amino acid

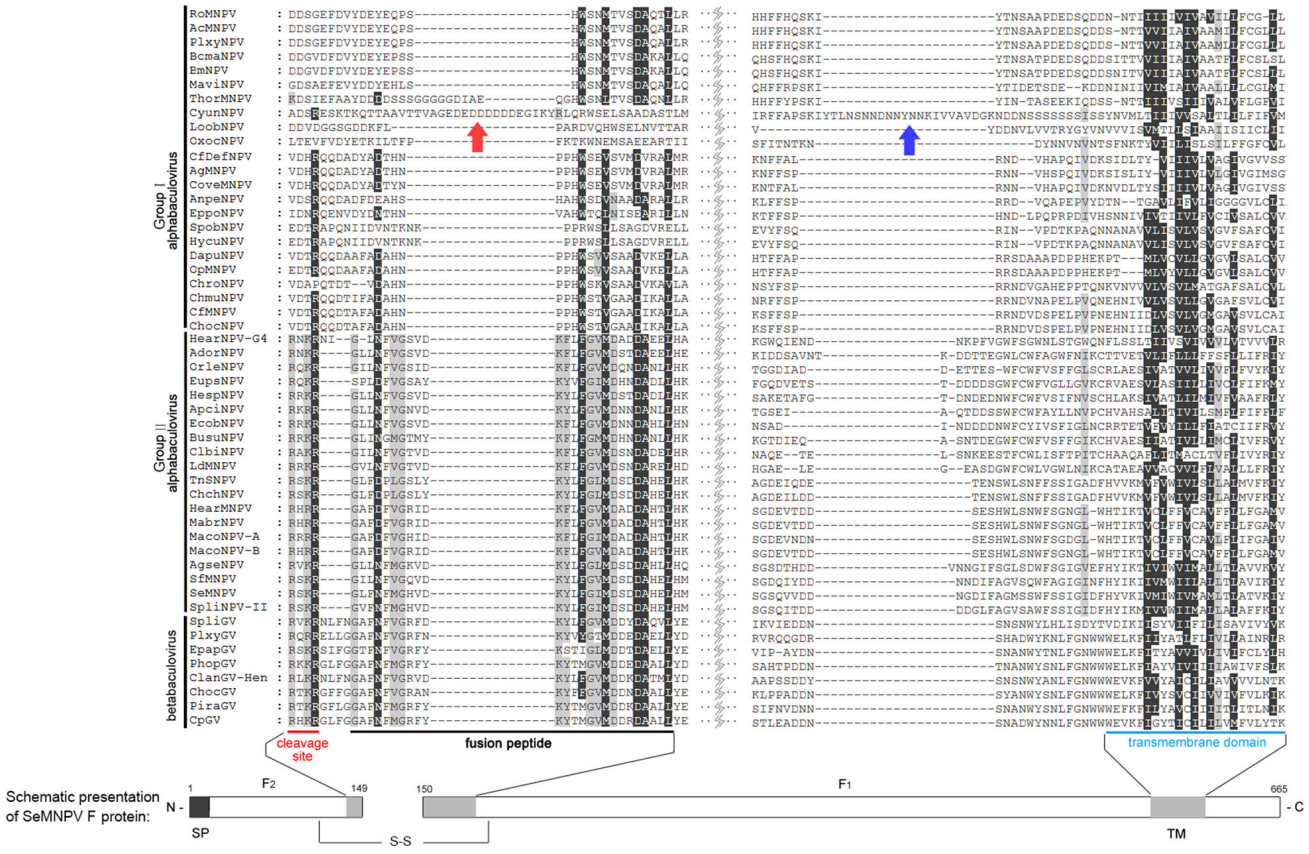

Fig. 5 Sequence alignment of F and F-like proteins. Alignment was performed using ClustalW. A schematic figure of SeMNPV F protein (Westenberg *et al.* 2004) with sequence alignments of two enlarged regions is shown at the bottom. The predicted regions of the furin cleavage site, fusion peptide, and TM domains are indicated below the alignments. The red and blue arrows indicate the insertion regions of the continuous polar amino acids in the fusion peptide region and pre-transmembrane domain, respectively in CyunNPV.

insertions interfere with hydrophobicity in these regions and are likely to affect fusogenicity. A similar insertion in the fusion peptide region has been identified in ThorNPV and could provide insights into the evolution of the F-like proteins. Our findings suggest that different Group I alphabaculoviruses underwent different routes to inactivate the fusion function of F targeting the furin cleavage site, the fusion peptide, or the preTM domain. The acquisition of higher fusogenicity by GP64 and the inactivation of the fusion capacity of F proteins are considered critical events in the emergence of Group I viruses. Our results provide new insights into the underlying F inactivation throughout evolution.

CyunNPV is a member of Group I alphabaculoviruses and exhibits certain unique features. Phylogenetic analysis shows that CyunNPV occupies a distinct branch in the alphabaculoviral clades. The F-like protein contains two insertions that are different from those found in other alphabaculoviruses and could reflect the process underlying the inactivation of the protein. The abovementioned unique features indicate that CyunNPV is a distinct species in the genus *Alphabaculovirus*.

Acknowledgements This work was supported by the National Key R&D Program of China (Grant No. 2017YFD0200400) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB11030400). We acknowledge the technical support and core facility of the Wuhan Institute of Virology for their technical assistance.

Author Contributions Conceived and designed the experiments: ZZ, MW, ZH, FD, HW, and FY. Viral nucleic acid extraction and 454 sequencing: ZZ, DH, FY, XL, and ZL. Genome assembly, annotation and data analysis: ZZ, XL, JW, QW, and HL. Wrote the paper: ZZ, BA, ZH, and MW.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Animal and Human Rights Statement This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Bulach DM, Kumar CA, Zaia A, Liang B, Tribe DE (1999) Group II nucleopolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedrin and DNA polymerase gene sequences. J Invertebr Pathol 73:59–73

Carstens EB, Ball LA (2009) Ratification vote on taxonomic proposals to the international committee on taxonomy of viruses (2008). Arch Virol 154:1181–1188

Conant GC, Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinformatics 24:861–862

Deng F, Wang R, Fang M, Jiang Y, Xu X, Wang H, Chen X, Arif BM, Guo L, Wang H, Hu Z (2007) Proteomics analysis of Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus identified two new occlusion-derived virus-associated proteins, HA44 and HA100. J Virol 81:9377–9385

Ferrelli ML, Salvador R, Biedma ME, Berretta MF, Haase S, Sciocco-Cap A, Ghiringhelli PD, Romanowski V (2012) Genome of Epinotia aporema granulovirus (EpapGV), a polyorganotropic fast killing betabaculovirus with a novel thymidylate kinase gene. BMC Genom 13:548

Garavaglia MJ, Miele SA, Iserte JA, Belaich MN, Ghiringhelli PD (2012) The ac53, ac78, ac101, and ac103 genes are newly discovered.core genes in the family Baculoviridae. J Virol 86:12069–12079

Gross CH, Shuman S (1998) Characterization of a baculovirus-encoded RNA 5'-triphosphatase. J Virol 72:7057–7063

Guarino LA, Gonzalez MA, Summers MD (1986) Complete sequence and enhancer function of the homologous DNA regions of Autographa californica nuclear polyhedrosis virus. J Virol 60:224–229

Habib S, Hasnain SE (2000) Differential activity of two non-hr origins during replication of the baculovirus Autographa californica nuclear polyhedrosis virus genome. J Virol 74:5182–5189

Hayakawa T, Ko R, Okano K, Seong SI, Goto C, Maeda S (1999) Sequence analysis of the Xestia c-nigrum granulovirus genome. Virology 262:277–297

Herniou EA, Olszewski JA, Cory JS, O'Reilly DR (2003) The genome sequence and evolution of baculoviruses. Annu Rev Entomol 48:211–234

Hu ZH, Arif BM, Jin F, Martens JW, Chen XW, Sun JS, Zuidema D, Goldbach RW, Vlak JM (1998a) Distinct gene arrangement in the Buzura suppressaria single-nucleocapsid nucleopolyhedrovirus genome. J Gen Virol 79:2841–2851

Hu ZH, Arif BM, Sun JS, Chen XW, Zuidema D, Goldbach RW, Vlak JM (1998b) Genetic organization of the HindIII-I region of the single-nucleocapsid nucleopolyhedrovirus of Buzura suppressaria. Virus Res 55:71–82

Huang JP, Tang WQ, Shu QH (1983) A study on cyclophragma undans nuclear polyhedrosis virus. J Cent South Univ For Technol 3:136–142 **(In Chinese)**

Huang J, Hao B, Deng F, Sun X, Wang H, Hu Z (2008) Open reading frame Bm21 of Bombyx mori nucleopolyhedrovirus is not essential for virus replication in vitro, but its deletion extends the median survival time of infected larvae. J Gen Virol 89:922–930

Ijkel WFJ, Westenberg M, Goldbach RW, Blissard GW, Vlak JM, Zuidema D (2000) A novel baculovirus envelope fusion protein with a proprotein convertase cleavage site. Virology 275:30–41

Jehle JA, Blissard GW, Bonning BC, Cory JS, Herniou EA, Rohrmann GF, Theilmann DA, Thiem SM, Vlak JM (2006) On the classification and nomenclature of baculoviruses: a proposal for revision. Arch Virol 151:1257–1266

Jiang Y, Deng F, Rayner S, Wang H, Hu Z (2009) Evidence of a major role of GP64 in group I alphabaculovirus evolution. Virus Res 142:85–91

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences Computer applications in the biosciences. CABIOS 8:275–282

Kamita SG, Nagasaka K, Chua JW, Shimada T, Mita K, Kobayashi M, Maeda S, Hammock BD (2005) A baculovirus-encoded protein tyrosine phosphatase gene induces enhanced locomotory activity in a lepidopteran host. Proc Natl Acad Sci USA 102:2584–2589

Kool M, Voeten JT, Goldbach RW, Tramper J, Vlak JM (1993) Identification of seven putative origins of Autographa californica

multiple nucleocapsid nuclear polyhedrosis virus DNA replication. J Gen Virol 74:2661–2668

Landais I, Vincent R, Bouton M, Devauchelle G, Duonor-Cerutti M, Ogliastro M (2006) Functional analysis of evolutionary conserved clustering of bZIP binding sites in the baculovirus homologous regions (hrs) suggests a cooperativity between host and viral transcription factors. Virology 344:421–431

Lauzon HA, Lucarotti CJ, Krell PJ, Feng Q, Retnakaran A, Arif BM (2004) Sequence and organization of the Neodiprion lecontei nucleopolyhedrovirus genome. J Virol 78:7023–7035

Li Y, Miller LK (1995) Properties of a baculovirus mutant defective in the protein phosphatase gene. J Virol 69:4533–4537

Liu X, Yin F, Zhu Z, Hou D, Wang J, Zhang L, Wang M, Wang H, Hu Z, Deng F (2014) Genomic sequencing and analysis of Sucra jujuba nucleopolyhedrovirus. PLoS ONE 9:e110023

Lorizate M, Huarte N, Saez-Cirion A, Nieva JL (2008) Interfacial pretransmembrane domains in viral proteins promoting membrane fusion and fission. Biochem Biophys Acta 1778:1624–1639

Lung OY, Cruz-Alvarez M, Blissard GW (2003) Ac23, an envelope fusion protein homolog in the baculovirus Autographa californica multicapsid nucleopolyhedrovirus, is a viral pathogenicity factor. J Virol 77:328–339

Pearson MN, Bjornson RM, Ahrens C, Rohrmann GF (1993) Identification and characterization of a putative origin of DNA replication in the genome of a baculovirus pathogenic for Orgyia pseudotsugata. Virology 197:715–725

Pearson MN, Groten C, Rohrmann GF (2000) Identification of the lymantria dispar nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the Baculoviridae. J Virol 74:6126–6131

Pearson MN, Russell RL, Rohrmann GF (2001) Characterization of a baculovirus-encoded protein that is associated with infected-cell membranes and budded virions. Virology 291:22–31

Rodems SM, Friesen PD (1995) Transcriptional enhancer activity of hr5 requires dual-palindrome half sites that mediate binding of a dimeric form of the baculovirus transregulator IE1. J Virol 69:5368–5375

Rohrmann GF, Karplus PA (2001) Relatedness of baculovirus and gypsy retrotransposon envelope proteins. BMC Evol Biol 1:1

Sainz B Jr, Rausch JM, Gallaher WR, Garry RF, Wimley WC (2005) The aromatic domain of the coronavirus class I viral fusion protein induces membrane permeabilization: putative role during viral entry. Biochemistry 44:947–958

Salzwedel K, West JT, Hunter E (1999) A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. J Virol 73:2469–2480

Sheng Z, Charbonneau H (1993) The baculovirus Autographa californica encodes a protein tyrosine phosphatase. J Biol Chem 268:4728–4733

Solovyev VV, Salamov AA (1999) INFOGENE: a database of known gene structures and predicted genes and proteins in sequences of genome sequencing projects. Nucleic Acids Res 27:248–250

Takagi T, Taylor GS, Kusakabe T, Charbonneau H, Buratowski S (1998) A protein tyrosine phosphatase-like protein from baculovirus has RNA 5′-triphosphatase and diphosphatase activities. Proc Natl Aca Sci USA 95:9808–9812

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Tan Y, Jiang L, Wang M, Yin F, Deng F, Liu M, Hu Z, Wang H (2008) Mutagenesis and nuclear magnetic resonance analyses of the fusion peptide of Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus F protein. J Virol 82:8138–8148

Theilmann DA, Stewart S (1992) Tandemly repeated sequence at the 3' end of the IE-2 gene of the baculovirus Orgyia pseudotsugata multicapsid nuclear polyhedrosis virus is an enhancer element. Virology 187:97–106

van Oers MM, Vlak JM (2007) Baculovirus genomics. Curr Drug Targets 8:1051–1068

Wang M, Tan Y, Yin F, Deng F, Vlak JM, Hu Z, Wang H (2008) The F-like protein Ac23 enhances the infectivity of the budded virus of gp64-null Autographa californica multinucleocapsid nucleopolyhedrovirus pseudotyped with baculovirus envelope fusion protein F. J Virol 82:9800–9804

Wang M, Wang J, Yin F, Tan Y, Deng F, Chen X, Jehle JA, Vlak JM, Hu Z, Wang H (2014) Unraveling the entry mechanism of baculoviruses and its evolutionary implications. J Virol 88:2301–2311

Westenberg M, Wang H, Ijkel WFJ, Goldbach RW, Vlak JM, Zuidema D (2002) Furin is involved in baculovirus envelope fusion protein activation. J Virol 76:178–184

Westenberg M, Veenman F, Roode EC, Goldbach RW, Vlak JM, Zuidema D (2004) Functional analysis of the putative fusion domain of the baculovirus envelope fusion protein F. J Virol 78:6946–6954

Wu Y, Carstens EB (1996) Initiation of baculovirus DNA replication: early promoter regions can function as infection-dependent replicating sequences in a plasmid-based replication assay. J Virol 70:6967–6972

Xiao KJ, Shu QH (1984) Biology of the oak caterpillar Cyclophragma undans (Walker). J Cent South Univ For Technol 4:59–64 **(In Chinese)**

Yin F, Zhu Z, Liu X, Hou D, Wang J, Zhang L, Wang M, Kou Z, Wang H, Deng F, Hu Z (2015) The complete genome of a new Betabaculovirus from Clostera anastomosis. PLoS ONE 10:e0132792

Yu IL, Lin YC, Robinson JH, Lung O (2009) Transduction of vertebrate cells with Spodoptera exigua multiple nucleopolyhedrovirus F protein-pseudotyped gp64-null Autographa californica multiple nucleopolyhedrovirus. J Gen Virol 90:2282–2287

Zanotto PM, Kessing BD, Maruniak JE (1993) Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. J Invertebr Pathol 62:147–164

Zhu Z, Yin F, Liu X, Hou D, Wang J, Zhang L, Arif B, Wang H, Deng F, Hu Z (2014) Genome sequence and analysis of Buzura suppressaria nucleopolyhedrovirus: a group II Alphabaculovirus. PLoS ONE 9:e86450