VIROLOGICA SINICA 2016, 31 (5): 425–436
DOI: 10.1007/s12250-016-3791-8

RESEARCH ARTICLE

De novo transcriptome analysis of *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV) genes in latently infected Se301 cells

Zheng Fang, Jingxu Shao, Qingbei Weng✉

School of Life Sciences, Guizhou Normal University, Guiyang 550001, China

Cells of the P8-Se301-C1 strain are *Spodoptera exigua* cell clones that each harbor a partial version of the *S. exigua* multiple nucleopolyhedrovirus (SeMNPV) genome and which are resistant to homologous SeMNPV infections. The cells produce no viral progeny, suggesting that the infection is a latent-like viral infection. To investigate the SeMNPV genes harbored in the P8-Se301-C1 cells, the *de novo* transcriptomes of P8-Se301-C1 cells and *S. exigua* Se301 cells were analyzed and compared. A total of 54,569,296 reads were obtained from the P8-Se301-C1 cells that yielded 112,565 final unigenes with a mean length of 1,093 nt. A total of 56,865,504 reads were obtained from the Se301 cells that yielded 102,996 final unigenes with a mean length of 1,082 nt. Ten SeMNPV gene transcripts (*se5*, *se7*, *se8*, *se12*, *se43*, *se45*, *se89*, *se90*, *se124*, and *se126*) were detected in the P8-Se301-C1 cells by RNA-Seq but not in the Se301 cells, which was verified by RT-PCR. 5'/3' RACE analyses showed that the 3'- or 5'-end sequences of the viral transcripts are aligned to the host gene sequences in P8-Se301-C1 cells, suggesting that the SeMNPV genes may integrate into and be transcribed with the host genes in the P8-Se301-C1 cells. Furthermore, six additional viral gene transcripts, *se11*, *se42*, *se44*, *se88*, *se91*, and *se127* (incorporated into chimeric fusion transcripts in the P8-Se301-C1 cells), were detected in the RACE analyses. Taken together, sixteen SeMNPV transcripts were identified in the P8-Se301-C1 cell strain. This study provides information to develop the understanding of baculovirus latent infections and superinfection exclusion.

KEYWORDS RNA-Seq; SeMNPV; baculovirus; latent infection; *Spodoptera exigua*

INTRODUCTION

The beet armyworm–*Spodoptera exigua*–is a major migratory insect pest that damages numerous vegetables and is causing increasing economic losses in the agriculture sector, including the industries related to cotton, food crops, and timber (Zheng et al., 2012; Virto et al., 2014; Qiu et al., 2015; Sun et al., 2015). The baculovirus *S. exigua* multiple nucleopolyhedrovirus (SeMNPV) is a very specific pathogen of *S. exigua*, hence, it has been developed for use as a bioinsecticide (Virto et al., 2014). Several bioinsecticides, including SPEXIT® (Andermatt Biocontrol, Switzerland), VIR-EX® (Biocolor, Spain), and SPOD-X® (Certis, US) contain SeMNPV.

Viral infections can be divided into acute, persistent, and latent infections. The vast majority viral genes are expressed during the acute infection and, because of the production of progeny virions, the infection spreads within a host and to new hosts (Saffert and Kalejta, 2007). During persistent infections, some viral genes are downregulated by viral or cellular regulatory gene products (Mayer and Ebbesen, 1994). A latent infection is defined as a reversible nonproductive infection of a cell in which the viral genome is present but infectious

Received: 22 April 2016, Accepted: 22 September 2016,
Published online: 18 October 2016
✉Correspondence:
Phone: +86-851-86702541, Fax: +86-851-86702541,
Email: wengqb@126.com
ORCID: 0000-0003-4244-573X

viruses are not produced except during intermittent episodes of reactivation (Stevens, 1989). During latent infections, viral genomes are maintained in host cells for a long periods with very little or no gene expression, which allows the virus to evade detection by the host immune system (Murillo et al., 2011).

P8-Se301 cells are *S. exigua* Se301 cells that are infected with an attenuated version of SeMNPV and the P8-Se301-C1 cell strain is cloned from these persistently infected P8-Se301 cells. P8-Se301-C1 cells harbor a partial SeMNPV genome and they are morphologically similar to Se301 cells but they do not produce viral progeny. The cells are resistant to SeMNPV infection but not to infection by the heterologous *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV), which is also a baculovirus. It has been suggested that SeMNPV resides in P8-Se301-C1 cells as a latent-like infection, which means that these cells provide a promising experimental system to investigate the mechanisms of baculovirus persistence in insects (Weng et al., 2009).

Herpes simplex virus 1 (HSV-1) can establish lifelong latency in the trigeminal sensory neurons of humans, and the expression of the viral RNA known as latency-associated transcript (LAT) in the absence of the production of viral proteins is believed to play a role in establishing latency (Perng et al., 1996; Thompson and sawtell, 1997). Therefore, exploring the transcription of SeMNPV genes in P8-Se301-C1 cells may aid the understanding of the molecular mechanism of latent infections.

Recently, Illumina strand-specific RNA sequencing (RNA-Seq)–a newly developed, large-scale, and genome-wide process–has been used for transcriptome analysis and gene discovery. This highlights the potential to use RNA-Seq to cost-effectively obtain large amounts of transcriptome data and then compare the evolution of the genomes of non-model species (Zhao et al., 2014; Slavokhotova et al., 2015; Xu et al., 2015). RNA-Seq has several obvious advantages, such as its cost-effectiveness, its high resolution, and the fact that it has a wide dynamic range of expression levels over which transcripts can be detected (Vogel et al., 2014; Lambirth et al., 2015). In this study, Illumina paired-end sequencing was used to analyze the *de novo* transcriptomes of *S. exigua* cells. We compared the transcriptome sequences of P8-Se301-C1 cells and Se301 cells using RNA-Seq and identified ten SeMNPV gene transcripts in the P8-Se301-C1 cells. Moreover, six additional SeMNPV gene transcripts were detected in P8-Se301-C1 cells using 5'/3' rapid amplification of cDNA ends (RACE) analyses. These novel findings provide useful information on the mechanisms of latent infections and superinfection exclusion.

MATERIALS AND METHODS

Cell culture

Se301 and P8-Se301-C1 cells were cultured at 27 °C in Grace's Insect Medium (Invitrogen, Carlsbad, US) supplemented with 10% fetal bovine serum (FBS), penicillin, (100 μg/mL) and streptomycin (30 μg/mL).

RNA extraction, cDNA synthesis, and Illumina sequencing

The total RNA was extracted from the two cell lines using a TaKaRa MiniBEST Universal RNA Extraction Kit (TaKaRa, Dalian, China) according to manufacturer's protocol. The RNA integrity number (RIN) of the total RNA was verified using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, US), and the quantity was measured using a NanoDrop 2000 Spectrophotometer.

The total RNA was purified further using an RNeasy Micro Kit (QIAGEN GmBH, Germany) and an RNase-Free DNase Set (QIAGEN). After removing the ribosomal RNA, the remaining RNA was split into short fragments using an RNA fragmentation buffer. The RNA fragments were used as templates to amplify first-strand cDNA using random hexamerprimers, and then the second cDNA strand was synthesized. cDNA libraries of both the Se301 cells and the P8-Se301-C1 cells were created using the double-stranded cDNA. Paired-end sequencing was carried out using the PE125 strategy of the Illumina HiSeq 2500 Sequencing System (Illumina, San Diego, CA, US) at Shanghai Biotechnology Corporation.

Sequence statistics, *de novo* assembly, and mapping

Before assembly and mapping, the raw RNA-Seq reads obtained from the Se301 and P8-Se301-C1 cDNA libraries were processed using the ShortRead package to filter out low-quality nucleotide sequences, adapters, and PCR primer sequences. Reads with a length shorter than 35 nt or with at least 2 ambiguous nucleotides (i.e., those which could be any type of nucleotide) were removed. The resulting cleaned reads were mapped to the SeMNPV genome to screen out the SeMNPV gene transcripts. The cleaned reads were assembled as primary unigenes using the Trinity package with an optimized k-mer length of 25 (Tao et al., 2012; Chen et al., 2014). The primary unigenes were cleaned by removing redundant genes and they were then assembled into a final set of unigenes using CD-HIT software (Yang and smith, 2013). The cleaned reads were aligned to the final unigenes using the mapping algorithm, FANSe2, and allowing up to 7 mismatched nucleotides (Zhang et al., 2012; Xiao et al., 2014). The final set of unigenes (with at least 10 mapped reads) were considered to be reliably assembled unigenes.

Annotation

Universal Protein (UniProt) is the most comprehensive catalog of protein sequences and functional annotations (Uniprot consortiums, 2011). In addition, the Gene Ontology (GO) database, the Clusters of Orthologous Groups of proteins (COG) database, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database are major databases of putative gene functions (He et al., 2012). BLASTX was used to search the UniProt, GO, COG, and KEGG databases for the final unigenes, with an e-value cut-off of 1e-5.

Reverse transcription PCR (RT-PCR)

RT-PCR was used to verify the SeMNPV gene transcripts from the P8-Se301-C1 cells. A cDNA First Strand Synthesis Kit (SiDanSai, Shanghai, China) was used to reverse transcribe 1 μg of total RNA extracted from the P8-Se301-C1 cells and from the Se301 cells into cDNA. The transcripts of the SeMNPV genes that were detected in the P8-Se301-C1 cells by RNA-Seq were amplified using RT-PCR and Se301 cells were used as a negative control. The specific primers are shown in Supplementary Table 1. Agarose gel electrophoresis was carried out to separate the resulting DNA products, which were then sequenced by Beijing Ruibo (Guangzhou, China).

5'/3' RACE analysis

To sequence the full-length of the SeMNPV gene transcripts, the total RNA was extracted from the P8-Se301-C1 cells and RACE analyses were carried out with a SMARTer RACE 5'/3' Kit (Clontech, US). A set of gene-specific primers were designed (Supplementary Table 1) and labelled either GSP (gene-specific primer) or NGSP (nested gene-specific primer) according to the sequence of the RT-PCR products. PCR amplification was carried using the following protocol. Briefly, reaction volumes of 20 μL were prepared containing 5'/3'-RACE-Ready cDNA as the template, primers (0.5 μL of the 10 μmol/L GSP and 2.0 μL of the universal primer mix provided in the kit) and 2×PrimerSTAR Max DNA Polymerase (TaKaRa). The conditions were as follows: 2 min at 94 °C; 25 cycles of 30 s at 94 °C, 30 s at 68 °C, 3 min at 72 °C; and a final extension at 72 °C for 5 min. The products were diluted with water (using a ratio of 1:10) and used as template cDNA for the nested PCR. The nested PCR conditions were as follows: 2 min at 94 °C; 30 cycles of 30 s at 94 °C, 30 s at 55 °C–60 °C, 3 min at 72 °C; and a final extension at 72 °C for 5 min. The PCR products were visualized by electrophoresis on a 1% agarose gel and sequenced. Subsequently, the 5' and 3' fragment sequences were assembled to obtain the full-length cDNA sequences of each of the SeMNPV genes.

RESULTS

Sequence trimming

A total of 56,865,504 raw reads were obtained from the Se301 cells and 54,569,296 were obtained from the P8-Se301-C1 cells. After a series of stringent filtering processes, 55,033,958 and 52,784,058 clean reads were obtained from the Se301 and P8-Se301-C1 cells, respectively (Supplementary Table S2).

De novo assembly

After primary assembly, we obtained 116,048 counts with a mean length of 1,173 nt and an N50 of 2,137 nt from the P8-Se301-C1 cells and 104,600 counts with a mean length of 1,169 nt and an N50 of 2,145 nt from the Se301 cells (Supplementary Table S3). After further assembly, we obtained 112,565 counts with a mean length of 1,093 nt and an N50 of 1,824 nt from the P8-Se301-C1 cells and 102,996 counts with a mean length of 1,082 nt and an N50 of 1,803 nt from the Se301 cells (Supplementary Table S4). The final set of unigenes from the P8-Se301-C1 cells comprised 24,731 unigenes (24.33%) that were ≥ 1,000 nt long and 10,229 unigenes (10.06%) that were > 2,000 nt long. The final set of unigenes from the Se301 cells comprised 29,849 unigenes (28.98%) that were ≥ 1,000 nt long and 13,949 unigenes (13.54%) that were > 2,000 nt long. The distributions of the lengths of the final unigenes from the P8-Se301-C1 and Se301 cells are presented in Supplementary Figure S1.

Functional annotation

Of the 112,565 final unigenes from the P8-Se301-C1 cells, 26,553 (23.6%) were annotated, and of the 102,996 final unigenes from the Se301 cells, 24,906 (24.2%) were annotated. The distributions of the BLASTX search (Uniprot database was used) hits for the P8-Se301-C1 and Se301 samples are shown in Figure 1. There were a significant number of hits for the unigenes from both cell lines with species in the order *Lepidoptera*.

The main GO categories were cellular components, biological processes, and molecular functions. The terms associated with the P8-Se301-C1 sample were all from these three main GO categories and comprised 53 functional subcategories) (Supplementary Figure S2). Regarding the cellular components category, the largest proportion of unigenes were assigned to the following subcategories: cells (20.52%), cell part (20.52%), membrane (14.28%) and membrane part (11.48%). The majority of the unigenes in the biological process category were assigned to the metabolic process subcategory (26.57%) and the cellular process subcategory (23.75%). Most of the unigenes in the molecular function category were related to binding (45.90%) and catalytic activity (38.61%), which included genes that encoded kinases,

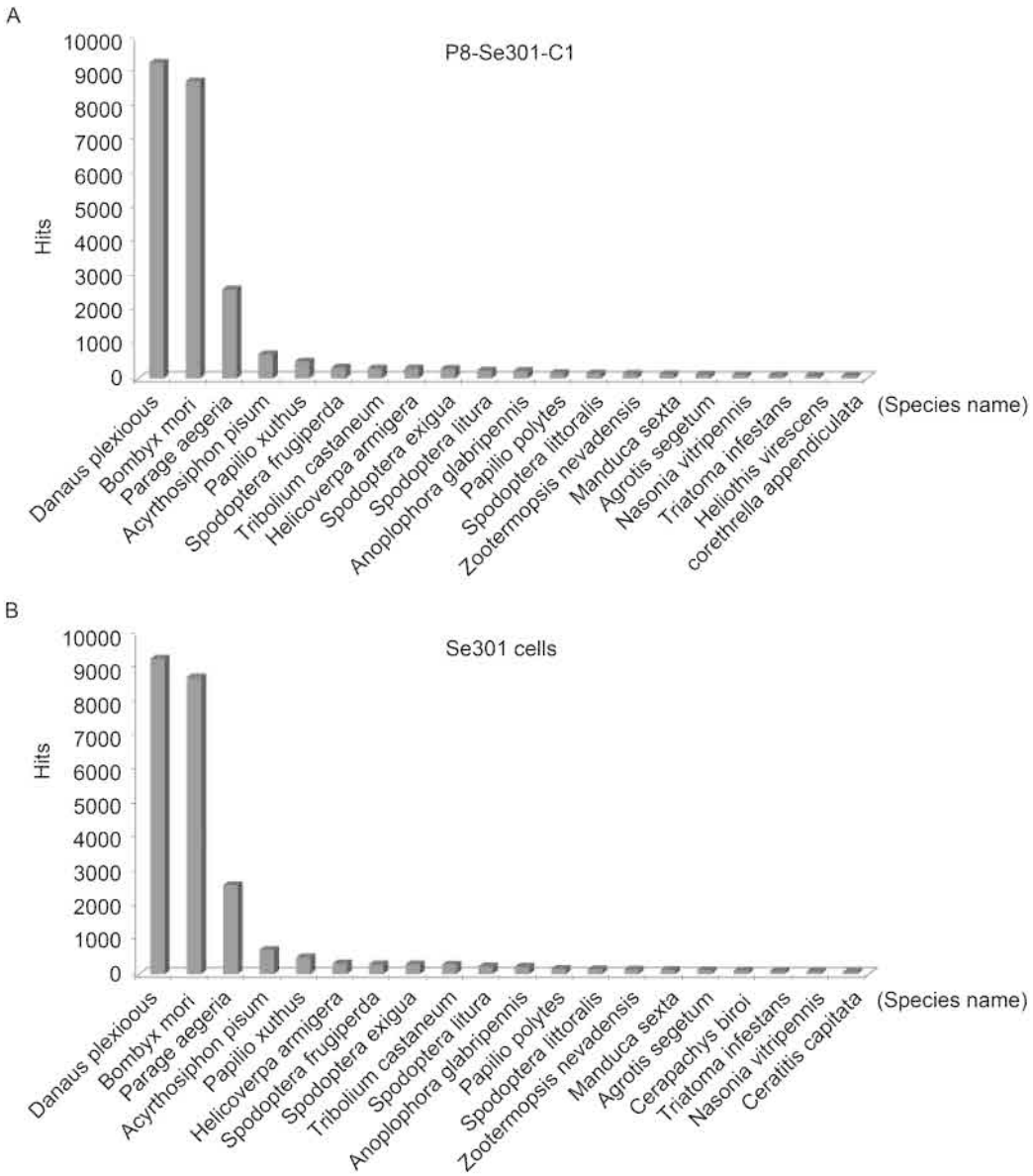


Figure 1. Distribution of the most frequent BLASTX hits (associated with 20 species) for the unigenes from the P8-Se301-C1 (A) and Se301 cells (B). There were a significant number of hits for the unigenes from both cell lines with species in the order *Lepidoptera*.

transferases, and hydrolases, many of which are likely to be involved in DNA replication, transcription, and translation.

To assist with the functional classification of the final set of unigenes, information on the functional classification of their homologs in the COG database was explored (Liu et al., 2016). A total of 19,825 unigenes from the P8-Se301-C1 cells were clustered into 25 COGs (Figure 2). Among them, the signal transduction mechanisms cluster was the largest (8.71%), followed by the general function prediction only cluster (6.75%). The other large clusters were transcription (5.03%), RNA processing and modification (4.48%), posttranslational modification, protein turnover and chaperones (4.40%), cytoskeleton (3.53%) and intracellular trafficking, secretion and vesicular transport (3.01%).

To identify the biological pathways that are active in the *S. exigua* cell lines, we mapped the 26,553 annotated sequences from P8-Se301-C1 cells to those associated with the canonical reference pathways in the KEGG database. In total, 26,553 unigenes were assigned to 290 known metabolic or signaling KEGG pathways. The top 11 KEGG pathways were metabolic pathways (2041 unigenes), biosynthesis of antibiotics (839), ribosomes

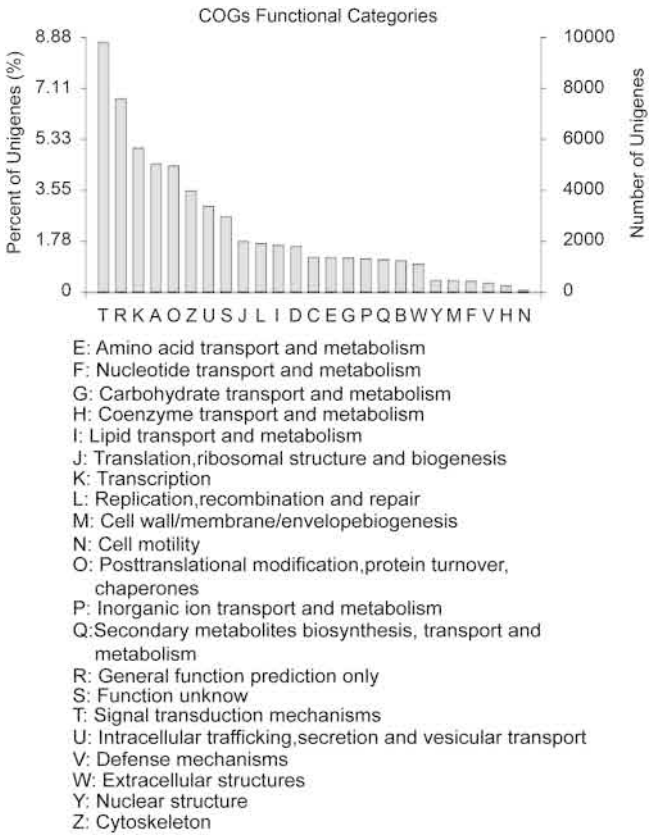


Figure 2. Clusters of orthologous groups (COG) classification of unigenes in P8-Se301-C1 sample.

(697), biosynthesis of secondary metabolites (687), protein processing in the endoplasmic reticulum (540), microbial metabolism in diverse environments (324), purine metabolism (308), HTLV-I infection (304), the PI3K-Akt signaling pathway (283), spliceosomes (279), and RNA transport (263).

Identification of SeMNPV gene transcripts in P8-Se301-C1 cells by RNA-Seq

After the clean reads from RNA-Seq were aligned with the SeMNPV genome, it was found that ten SeMNPV gene transcripts from the P8-Se301-C1 cells were not identified in the Se301 samples: *se5* (unknown function), *se7* (*me53*), *se8* (envelope fusion protein), *se12* (*lef2*), *se43* (unknown function), *se45* (ribonucleotide reductase small subunit), *se89* (unknown function), *se90* (unknown function), *se124* and *se126* (DNA binding protein).

RT-PCR was carried out to confirm the presence of the ten SeMNPV gene transcripts in the P8-Se301-C1 cells. The RT-PCR products were distinguished by agarose gel electrophoresis, and fragments of the expected sizes were detected for the P8-Se301-C1 samples, but not for the Se301 samples (Figure 3). The RT-PCR products were further verified by sequencing and by using a BLAST search of the NCBI's Nucleotide Database. All the RNA samples prepared in this study were subjected to RT-PCR with the same primers and the results indicated that there was no DNA contamination.

3′/5′ RACE analyses of the SeMNPV gene transcripts

To further understand the full-length transcripts of the ten SeMNPV genes from the P8-Se301-C1 cells, both 3' and 5' RACE analyses were performed using the total RNA extracts. No gene products were detected in the first round of amplification. However, after the second round of nested PCR using the RACE amplification products, a range of fragment products were observed (Figure 4).

The fragments of the 3′ and 5′ ends of the gene transcripts were recovered and sequenced, and then the sequences were assembled (along with the sequences from the previous RNA-Seq) into full-length transcripts. The full-length sequences of the transcripts were used to carry out a BLAST search of the NCBI's Nucleotide Database, and six additional SeMNPV gene transcripts from the P8-Se301-C1 cells were identified, including *se11* (*orf4 PE*), *se42* (unknown function), *se44* (unknown function), *se88* (*iap-2*), *se91* (*lef-3*) and *se127* (*lef-6*) (Figure 5A and Supplementary Figure S3). Thus, in total, sixteen viral gene transcripts that mapped to the SeMNPV genome were identified in the P8-Se301-C1 cells by a combination of RNA-Seq and RACE (Figure 5A), and the genes were found to be located in a disperse region of the SeMNPV genome (Figure 5B).

The full-length sequences of the transcripts showed that the SeMNPV genes are incorporated into the host cell genome. The genes consequently form chimeric fusion transcripts in P8-Se301-C1 cells, and either the 3' or 5' end of each transcript is aligned with the host genome (Figure 5A and Supplementary Figure S3).

The organization of each fusion transcript is as follows (Table 1). (1) The full-length transcript containing the *se5* gene maps to the nt positions in the SeMNPV genome at 6190–7713, the 5' end sequence (nt 3–48) aligns to the *Bombyx mori akh2* mRNA (nt 26–71) that encodes adipokinetic hormone-2 and the 3' end sequence cannot be aligned to any known sequence in the NCBI's Nucleotide Database. (2) The full-length transcript containing *se7* maps to nt 9261–10433 of the SeMNPV genome, the 5' end sequence cannot be aligned to any known sequence and the 3' end sequence (nt 1373–1448) aligns to the *B. mori dh40* mRNA (nt 3116–3192) that encodes diuretic hormone 40. (3) The full-length transcript containing *se8* maps to nt 12498–14495 of SeMNPV the genome) and the 5' end sequence (nt 12–60) aligns to the *B. mori dh40* mRNA (nt 1035–998). (4) The full-length transcript containing *se12* (which map to nt 16064–

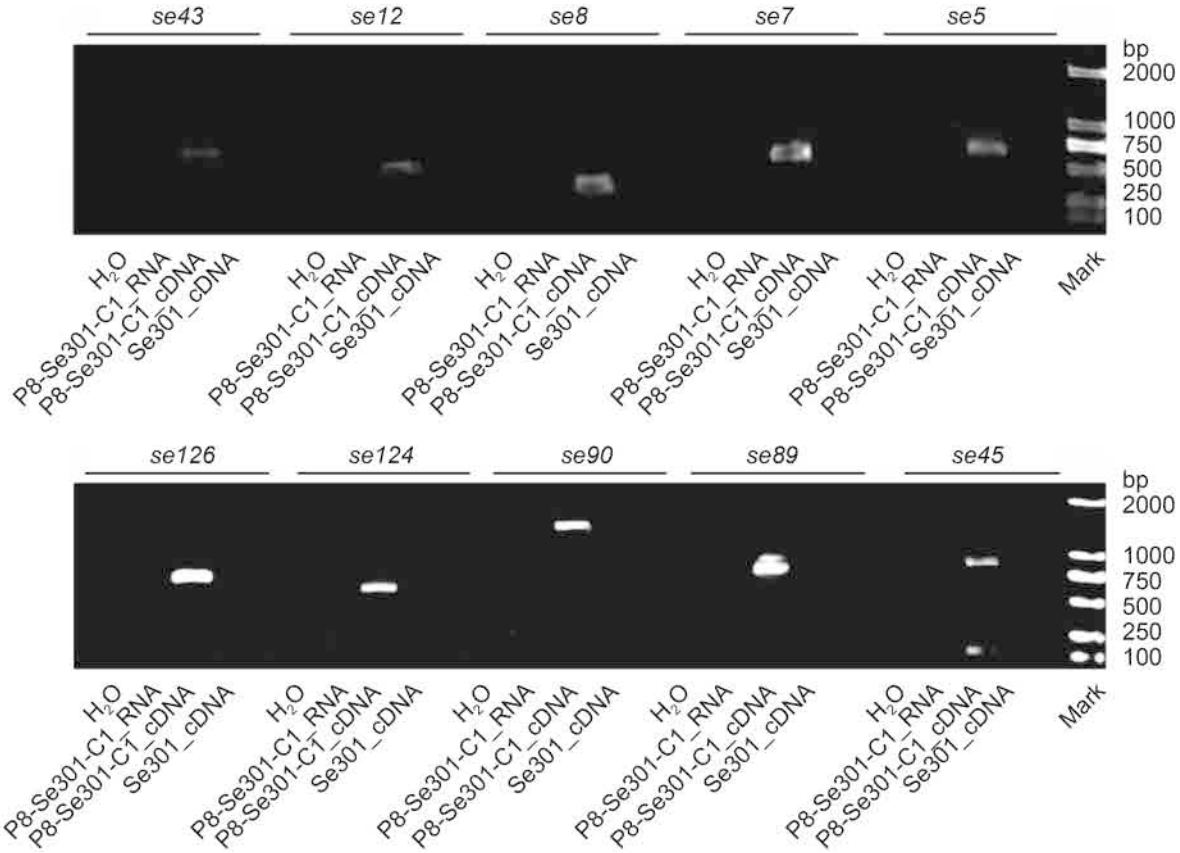


Figure 3. Ten SeMNPV genes (from the RNA-Seq analysis) verified by RT-PCR. The total RNA was extracted from the P8-Se301-C1 and Se301 cells and cDNAs were constructed. Subsequently, the ten SeMNPV genes were amplified by PCR and visualized on 1% agarose gel (the names of each of the genes are shown above each lane). PCRs were also performed with the RNA samples in place of the cDNA to exclude the possibility of DNA contamination.

16693 of SeMNPV genome) and the partial sequences of *se11* (which map to nt 15817–15852 and nt 15923–16101 of the SeMNPV genome) are transcribed together. The 5' end sequence (nt 5–48) aligns to the *B. mori akh2* mRNA (nt 27–71) and the 3' end sequence cannot be aligned to any known sequence. (5) The full-length transcript containing *se43* (which maps to nt 42696–43856 of the SeMNPV genome) and the partial sequence of *se42* (which maps to nt 42392–42606 of the SeMNPV genome) are transcribed together. Neither the 5' end nor the 3' end can be aligned to any known sequence. (6) The full-length transcript containing *se45* (which maps to nt 44408–45394 of the SeMNPV genome) and the partial sequence of *se44* (which maps to nt 44305–44312 of the SeMNPV genome) are transcribed together. The 5' end sequence cannot be aligned to any known sequence. (7) The full-length transcript containing *se89* (which maps to nt 85499–86398 of the SeMNPV genome) and the partial sequence of *se88* (which maps to nt 85088–85799 of the SeMNPV genome) are transcribed together. The 5' end sequence cannot be aligned to any known sequence. (8) The full-length transcript containing *se124* (which maps to nt 118809–119391 of the SeMNPV genome), the partial sequences of *se90* (which maps to nt 86521–86789 of the SeMNPV genome) and *se91* (which maps to nt 86788–86860 of the SeMNPV genome) are transcribed together. The 5' sequence (nt 3–50) aligns to the *B. mori dh40* mRNA (nt 25–71). (9) The full-length transcript containing *se126* (which maps to nt 120802–121788 of the SeMNPV genome) and the partial sequence of *se127* (which maps to nt 121816–122307 of the SeMNPV genome) are transcribed together. The 5' end sequence cannot be aligned to any known sequence and the 3' end sequence (nt 1360–1413) aligns to the *B. mori sifa* mRNA (nt 462–564) that encodes SIFamide.

DISCUSSION

Although bioinformatics tools for sequence assembly and data analysis have been developed, the *de novo* assembly of short reads continues to be challenging in cases where there is no reference genome (Chen et al.,

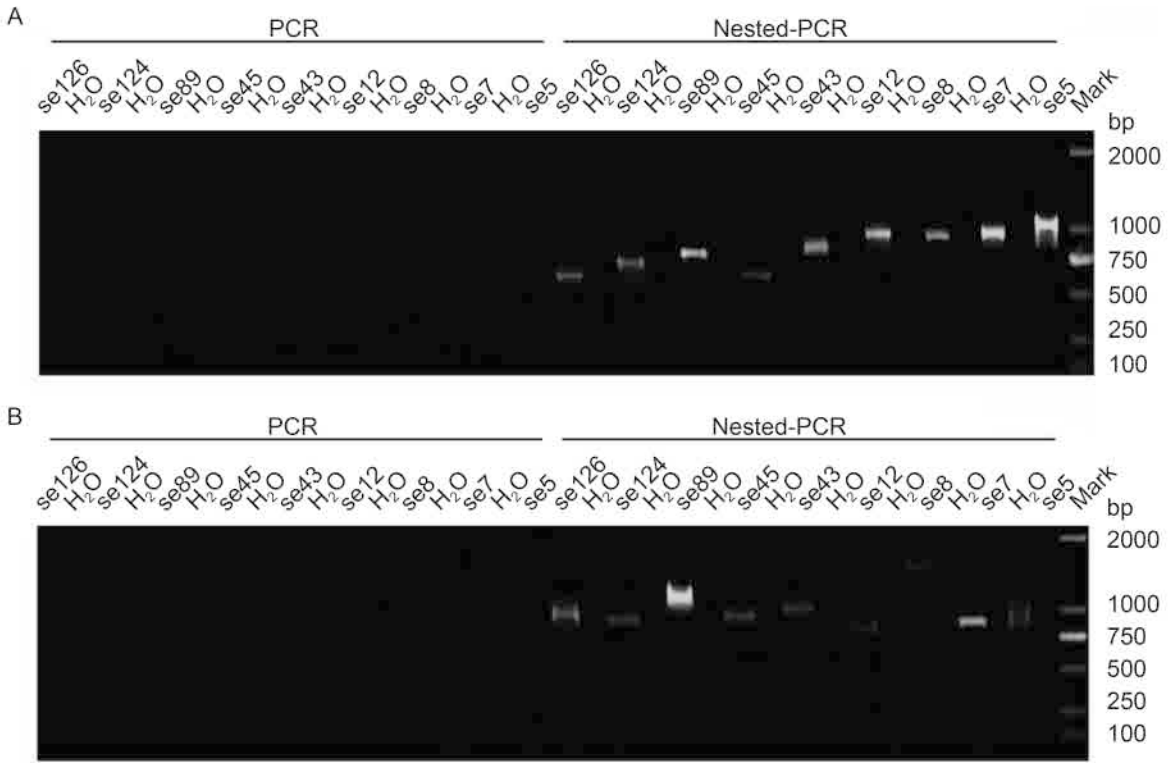


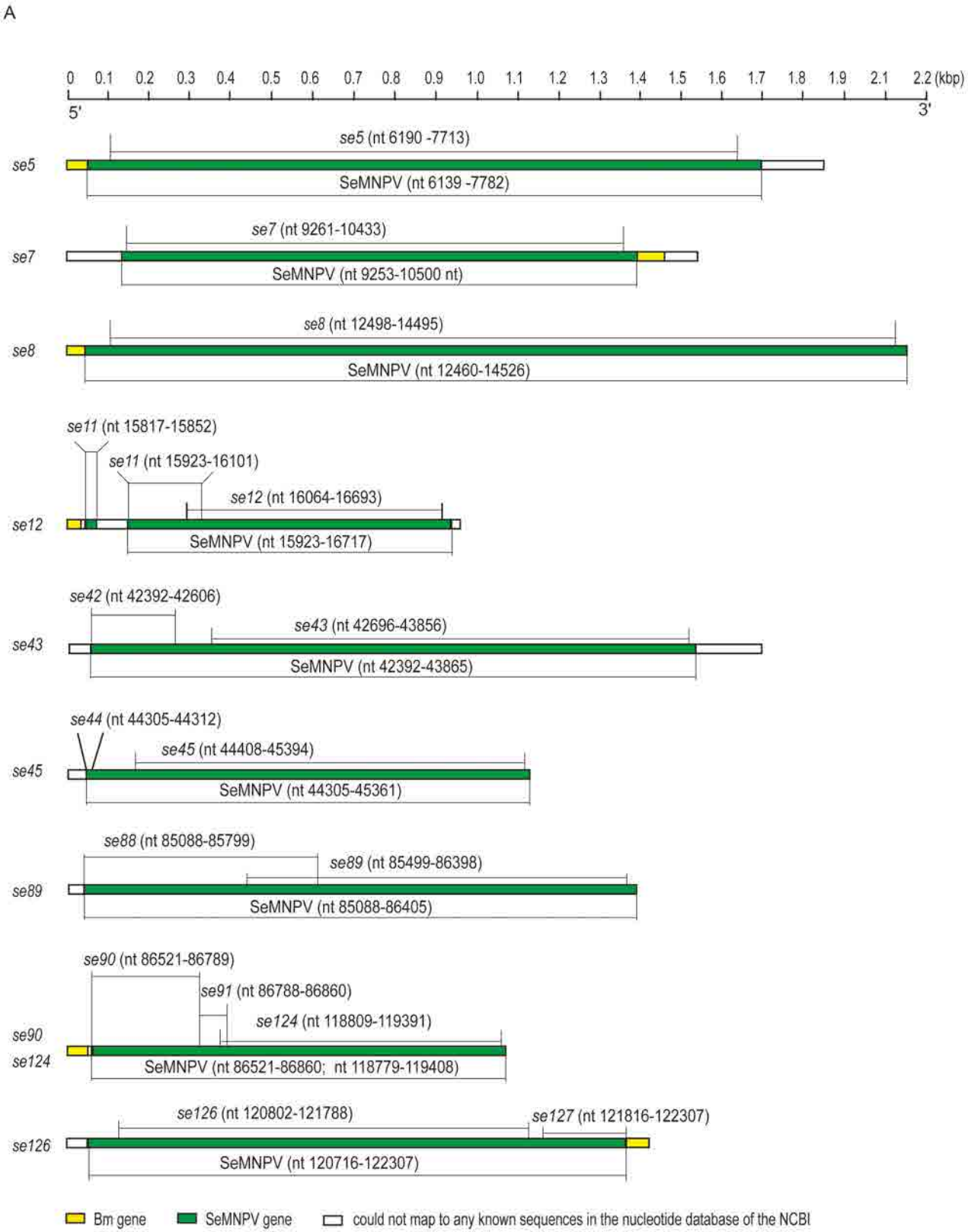
Figure 4. RACE analyses of the 3' and 5' end sequences of the SeMNPV gene transcripts in P8-Se301-C1 cells. (A) 3′ end RACE analysis. (B) 5′ end RACE analysis. The total RNA was extracted from the P8-Se301-C1 cells and 5'/3' RACE analyses were performed to determine the nucleotide sequences of the 5' and 3' ends of the SeMNPV gene transcripts. The PCR products were visualized on 1% agarose gel. No gene products were detected in the first round of amplification. However, after a second round of nested PCR on the RACE amplification products, a range of fragment products were observed. H_2O were used to as a control.

2014). In this study, we obtained 56,865,504 raw reads from Se301 cells and 54,569,296 raw reads from P8-Se301-C1 cells using RNA-Seq. Previous studies have obtained more than 34,809,334 raw reads associated with the *S. exigua* transcriptome and representing all the stages of the lifecycle of *S. exigua* comprising the egg, larval, pupal, and adult stages (Li et al., 2013) and 74,928 raw reads associated with the gene expression of *S. exigua* larvae infected with AcMNPV (Choi et al., 2012). After assembly of the *de novo* transcripts from the P8-Se301-C1 and Se301 samples using Trinity software, a total of 112,565 and 102,996 final unigenes were generated, respectively. The set of final unigenes from the P8-Se301-C1 cells contained 26,553 (23.6%) unigenes that were annotated and the final set of unigenes from the Se301 cells contained 24,906 (24.2%) unigenes that were annotated. This additional data provides more genetic information for further studies of *S. exigua*.

In the P8-Se301-C1 cells, ten SeMNPV gene transcripts were detected by RNA-Seq. RACE analyses were carried out to analyze the full length of each of the ten gene transcripts, and six additional SeMNPV genes were detected during the RACE analyses. Thus, taken together, sixteen SeMNPV gene transcripts were identified in P8-Se301-C1 cells, comprising the full-length sequences of nine genes (*se5*, *se7*, *se8*, *se12*, *se43*, *se45*, *se89*, *se124* and *se126*), and the partial sequences of seven genes (*se11*, *se42*, *se44*, *se88*, *se90*, *se91* and *se127*).

se5 is a unique SeMNPV gene with unknown function. A bioinformatics analysis showed that it has five protein kinase C phosphorylation sites and nine casein kinase II phosphorylation sites (Sun et al., 2015). Compared to the wild-type control, the deletion of *se5* resulted in a decrease in the pathogenicity of occlusion bodies (OBs) and a significantly extended the viral life cycle. *se7* is an early gene that is unnecessary for viral replication but required for efficient production of baculoviruses (de jong et al., 2009). *se8*, which encodes F protein, has been found in group II nucleopolyhedroviruses (NPVs). It is a functional homolog of the GP64 protein of group I NPVs of alphabaculoviruses but has a higher rate of activitythan GP64 (Westenberg et al., 2007).

se11 is a closely related homolog of *ld137a* (*orf4 PE*) from *Lymantria dispar* multiple nucleopolyhedrovirus (LdMNPV), which encodes a protein that appears to be a major component of the polyhedron envelope (Gombart


A
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.1 2.2 (kbp)
5'
3'
se5
se5 (nt 6190 -7713)
SeMNPV (nt 6139 -7782)
se7
se7 (nt 9261-10433)
SeMNPV (nt 9253-10500 nt)
se8
se8 (nt 12498-14495)
SeMNPV (nt 12460-14526)
se12
se11 (nt 15817-15852)
se11 (nt 15923-16101)
se12 (nt 16064-16693)
SeMNPV (nt 15923-16717)
se43
se42 (nt 42392-42606)
se43 (nt 42696-43856)
SeMNPV (nt 42392-43865)
se45
se44 (nt 44305-44312)
se45 (nt 44408-45394)
SeMNPV (nt 44305-45361)
se89
se88 (nt 85088-85799)
se89 (nt 85499-86398)
SeMNPV (nt 85088-86405)
se90
se124
se90 (nt 86521-86789)
se91 (nt 86788-86860)
se124 (nt 118809-119391)
SeMNPV (nt 86521-86860; nt 118779-119408)
se126
se126 (nt 120802-121788)
se127 (nt 121816-122307)
SeMNPV (nt 120716-122307)
Bm gene
SeMNPV gene
could not map to any known sequences in the nucleotide database of the NCBI

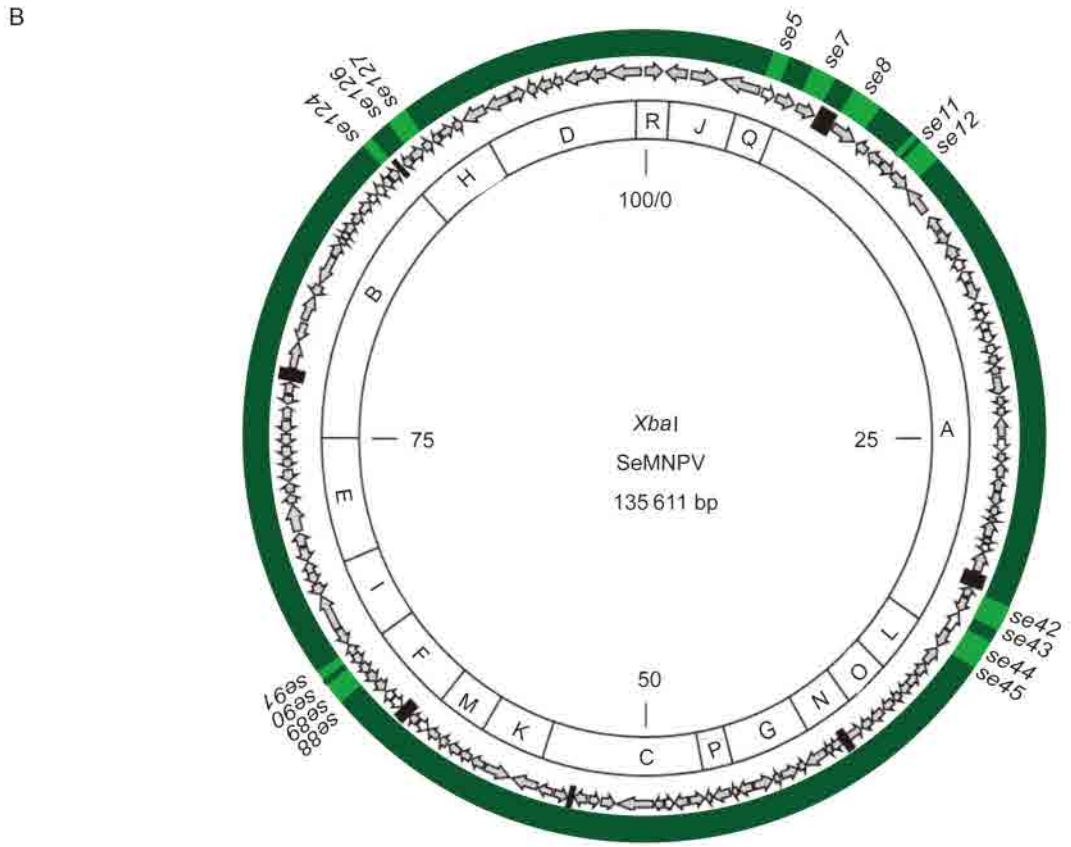


Figure 5. (A) Chimeric SeMNPV gene transcripts from the P8-Se301-C1 cells. The fragments of the 3' and 5' ends of the ten SeMNPV gene transcripts detected by RNA-Seq (shown on the left) were sequenced and assembled into full-length cDNA sequences. Six more SeMNPV gene transcripts were detected in RACE analyses. The positions of all sixteen detected SeMNPV genes (in green) are shown aligned to the SeMNPV genome. The SeMNPV genes integrate into the host genome and the 3′ or 5′ ends of the SeMNPV gene transcripts (in green) are aligned to the host genome (in yellow). The bar at the top of the figure is marked to show the size of each of the transcripts. (B) Overview of the SeMNPV transcripts from the P8-Se301-C1 cells. The sixteen SeMNPV genes (in green) detected by RNA-Seq and RACE analyses map to the SeMNPV genome (in dark green). The circular map has been modified according to the map of the SeMNPV genome (IJkel et al., 1999).

et al., 1989a; Russell & Rohrmann, 1990). *se12* is a late gene expression factor that plays a direct role in very late gene transcription (Sriram and Gopinathan, 1998). *se42* is the closely related homolog of *ac19* from AcMNPV (Ijkel et al., 1999). *se44* is a unique SeMNPV gene with unknown function (Ijkel et al., 1999). The homolog of *se43* in AcMNPV is *ac18*, which is a highly conserved gene in lepidopteran nucleopolyhedroviruses. *se43* may play a role in the efficiency of *S. exigua* infections (Wang et al., 2007).

In terms of protein identity and genomic location, the homolog of *se45* in LdMNPV is *ld120* (rr2b) (Ijkel et al., 1999). *se88* is a closely related homolog of *ac71* (*iap-2*) from AcMNPV, which functions as an apoptosis inhibitor and is a potential host range factor (Li et al., 1999). The homolog of *se89* in AcMNPV, *ac69*, has a stimulatory effect on late gene expression and has been implicated in cell division (Li et al., 1999). *se90* is a baculovirus core gene that has been shown to be highly conserved in all baculoviruses that have been fully sequenced (Nie et al., 2012). It is homologous to ORF 56 (*bm56*) of *B. mori* nuclear polyhedrosis virus (BmNPV), which facilitate efficient virus production *in vivo* (Xu et al., 2008). *se91* is a closely related homolog of *ac67* (*lef-3*) from AcMNPV, which is essential for AcMNPV DNA replication *in vivo* (Yu and carstens, 2012).

The *se124* homologs are highly conserved in all sequenced alphabaculoviruses and its homolog in AcMNPV, Ac34, is an activator protein that promotes late gene expression and is essential for the pathogenicity of AcMNPV (Cai et al., 2012). *se126* is a homolog of ORF16 (*bm16*) of BmNPV, which functions as a single-stranded DNA binding protein that plays a role in virus replication (Mikhailov et al., 1998). *se127* is a closely related homolog

Table 1. The SeMNPV gene-containing full-length transcripts in P8-Se301-C1 cells

Transcript size (bp)	Harbored SeMNPV genes		5' end sequence		3' end sequence	
	Position in the transcript	SeMNPV gene (position in the SeMNPV genome)	Position in the transcript	Aligned host gene mRNA (position in the host gene)	Position in the transcript	Aligned host gene mRNA (position in the gene)
1847	nt 106–1634	*se5* (nt 6190–7713)	nt 3–48	*B. mori akh2* (nt 26–71)	nt 1707–1847	No hit[a]
1530	nt 138-1310	*se7* (nt 9261–10433)	nt 1–130	No hit	nt 1373–1448	*B. mori dh40* (nt 3116–3192)
2136	nt 104–2103	*se8* (nt 12498–14495)	nt 12–60	*B. mori dh40* (nt 1035–998)	–[b]	–
978	nt 300–933 nt 55–91 nt 162-229	*se12* (nt 16064–16693) *se11* (nt 15817–15852) *se11* (nt 15923–16101)	nt 5–48	*B. mori akh2* (nt 27–71)	nt 946–978	No hit
1691	nt 348–1511 nt 46–261	*se43* (nt 42696–43856) *se42* (nt 42392–42606)	nt 1–46	No hit	nt 1520–1691	No hit
1118	nt 161–1107 nt 62–70	*se45* (nt 44408–45394) *se44* (nt 44305–44312)	nt 1–61	No hit	–	–
1370	nt 488–671 nt 35–489	*se89* (nt 85499–86398) *se88* (nt 85088–85799)	nt 1–35	No hit	–	–
1035	nt 407–991 nt 57–323 nt 322–396	*se124* (nt 118809–119391) *se90* (nt 86521–86789) *se91* (nt 86788–86860)	nt 3–50	*B. mori adh40* (nt 25–71)	nt 1007–1035	–
1421	nt 102–1122 nt 1164–1354	*se126* (nt 120802–121788) *se127* (nt 121816–122307)	nt 1–62	No hit	nt 1360–1413	*B. mori sifa* (nt 462–564)

Note: a: The end sequence cannot be aligned to any known sequence in the NCBI's Nucleotide Database. b: The end sequence is matched to SeMNPV gene.

of *ac28* (*lef-6*) of AcMNPV, which is not essential for viral replication but the infection cycle is substantially delayed in its absence (Lin and blissard, 2002).

Previously, a study demonstrated the presence of a SeMNPV *polyhedrin* transcript in P8-Se301-C1 cells (Weng et al., 2009). In this study, before RNA-Seq was carried out, the presence of the *polyhedrin* transcript was confirmed by nested RT-PCR using the total RNA extracted from the P8-Se301-C1 cells. However, the transcript was not identified from the RNA-Seq results. Six more transcripts (*se11*, *se42*, *se44*, *se88*, *se91* and *se127*) were detected using RACE analyses, but not by RNA-Seq, which implies that more SeMNPV gene transcripts may be present in P8-Se301-C1 cells but were not identified due to their low abundance.

The RACE analyses of the full-length transcripts containing SeMNPV genes showed that the 3′ and 5′ end sequences of the transcripts map to *B. mori* genes or to unknown genes that cannot be aligned to sequences in the NCBI's Nucleotide Database. We also attempted to align the 3′ and 5′ end sequences to the previous sequence data on *S. exigua* (NCBI accession numbers: SRX110132, SRX110248, GAFU00000000 and SRA056289) (Choi et al., 2012; Li et al., 2013), but there were no matches. As the entire SeMNPV genome has been sequenced (Ijkel et al., 1999), the 3′ and 5′ end must be from the host genome and not from the SeMNPV genome, suggesting that the SeMNPV genes integrate into host genome and are transcribed with the host genome in P8-Se301-C1 cells.

Latent infections of insect cells with nudivirus 1 (HzNV-1) is characterized by the expression of only one viral transcript (persistence associated transcript 1, PAT1) (Chao et al., 1992), and the viral DNA exists as either a circular or an inserted form (Lee and lehman, 1999). In chronic hepatitis B virus (HBV) infections of humans, subgenomic HBV DNA fragments were found to integrate into random sites of the host genome (Bonilla and Roberts, 2005; Minami et al., 2005). Due to the integration, the host genome was altered (a "cis" effect) and the HBV genome was altered (a "trans" effect) (Bonilla and Roberts, 2005; Minami et al., 2005), which may led to the modulation of the expression of the human genes

near the integration sites, followed by integration site-specific expression of the genes during hepatocarcinogenesis (Tamori et al., 2005). In our study, we found that SeMNPV genes were integrated into the host genome in the P8-Se301-C1 cells, which led to the production of novel fusion transcripts. The results support our previous findings (Weng et al., 2009) and further demonstrate that SeMNPV resides in P8-Se301-C1 cells as a latent infection.

In summary, we obtained fully assembled transcripts from *S. exigua* cells and identified sixteen transcripts of SeMNPV genes in P8-Se301-C1 cells. Among the SeMNPV gene transcripts, ten transcripts were detected using RNA-Seq and six additional transcripts were detected using RACE analyses. The findings suggest that, in Se301 cells that are latently infected with SeMNPV, some baculoviral genomic DNA fragments fuse together and integrate into the host genome, which leads to the production of novel chimeric fusion transcripts. The role of these novel fusion transcripts in latency in P8-Se301-C1 cells remains unclear. Future DNA sequencing studies would be useful to further explain the form of the viral DNA in the latently infected insect cells and the locations of the viral gene integration sites in the host genome. The findings provide important insights into the molecular mechanisms of baculovirus latent infections and superinfection exclusion.

ACKNOWLEDGMENTS

The authors appreciate valuable thoughts and suggestions from Prof. Kai Yang and Dr. Wenbi Wu at State Key Laboratory of Biocontrol of Sun Yet-sen University. We also thank Hanquan Liang at Sun Yet-sen University for revisions. This research was funded by the National Natural Science Foundation of China (No. 31160378), the Science and Technology Foundation of Guizhou Province (No. LKS[2010]21), and the Startup Foundation for Doctors of Guizhou Normal University.

COMPLIANCE WITH ETHICS GUIDELINES

The authors declared that they have no conflict of interest. This article does not contain any studies with human or animal subjects performed by any of the authors.

AUTHOR CONTRIBUTIONS

QBW conceived and supervised the study, and finalized the manuscript. ZF performed most experiments and wrote the paper. JXS helped with the RACE experiments and manuscript drafting. All authors read and approved the final manuscript.

Supplementary figures/tables are available on the websites of *Virologica Sinica*: www.virosin.org; link.springer.com/journal/12250.

REFERENCES

Bonilla GR, Roberts LR. 2005. The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. J Hepatol, 42: 760–777.

Cai Y, Long Z, Qiu J, Yuan M, Li G, Yang K. 2012. An ac34 deletion mutant of Autographa californica nucleopolyhedrovirus exhibits delayed late gene expression and a lack of virulence in vivo. J Virol, 86: 10432–10443.

Chao YC, Wood HA, Chang CY, Lee HJ, Shen WC, Lee HT. 1992. Differential expression of Hz-1 baculovirus genes during productive and persistent viral infections. J Virol, 66: 1442–1448.

Chen B, Zhang YJ, He Z, Li W, Si F, Tang Y, He Q, Qiao L, Yan Z, Fu W, Che Y. 2014. De novo transcriptome sequencing and sequence analysis of the malaria vector Anopheles sinensis (Diptera: Culicidae). Parasit Vectors, 7: 314.

Choi JY, Roh JY, Wang Y, Zhen Z, Tao XY, Lee JH, Liu Q, Kim JS, Shin SW, Je YH. 2012. Analysis of genes expression of Spodoptera exigua larvae upon AcMNPV infection. PLoS One, 7: e42462.

de Jong J, Arif BM, Theilmann DA, Krell PJ. 2009. Autographa californica multiple nucleopolyhedrovirus me53 (ac140) is a nonessential gene required for efficient budded-virus production. J Virol, 83: 7440–7448.

He W, You M, Vasseur L, Yang G, Xie M, Cui K, Bai J, Liu C, Li X, Xu X, Huang S. 2012. Developmental and insecticide-resistant insights from the de novo assembled transcriptome of the diamondback moth, Plutella xylostella. Genomics, 99: 169–177.

IJkel W, van Strien EA, Heldens JG, Broer R, Zuidema D, Goldbach RW, Vlak JM. 1999. Sequence and organization of the Spodoptera exigua multicapsid nucleopolyhedrovirus genome. J Gen Virol, 80: 3289–3304.

Lambirth KC, Whaley AM, Blakley IC, Schlueter JA, Bost KL, Loraine AE, Piller KJ. 2015. A Comparison of transgenic and wild type soybean seeds: analysis of transcriptome profiles using RNA-Seq. BMC Biotechnol, 15: 89.

Lee SS, Lehman IR. 1999. The interaction of herpes simplex type 1 virus origin-binding protein (UL9 protein) with Box I, the high affinity element of the viral origin of DNA replication. J Biol Chem, 274: 18613–18617.

Li L, Harwood SH, Rohrmann GF. 1999. Identification of additional genes that influence baculovirus late gene expression. Virology, 255: 9–19.

Li H, Jiang W, Zhang Z, Xing Y, Li F. 2013. Transcriptome Analysis and Screening for Potential Target Genes for RNAi-Mediated Pest Control of the Beet Armyworm, Spodoptera exigua. PLoS One, 8: e65931.

Lin G, Blissard GW. 2002. Analysis of an Autographa californica multicapsid nucleopolyhedrovirus lef-6-null virus: LEF-6 is not essential for viral replication but appears to accelerate late gene transcription. J Virol, 76: 5503–5514.

Liu H, Wu W, Hou K, Chen J, Zhao Z. 2016. Deep sequencing reveals transcriptome re-programming of Polygonum multiflorum thunb. roots to the elicitation with methyl jasmonate. Mol Genet Genomics, 291: 337–348.

Mikhailov VS, Mikhailova AL, Iwanaga M, Gomi S, Maeda S. 1998. Bombyx mori nucleopolyhedrovirus encodes a DNA-binding protein capable of destabilizing duplex DNA. J Virol, 72: 3107–3116.

Minami M, Daimon Y, Mori K, Takashima H, Nakajima T, Itoh Y, Okanoue T. 2005. Hepatitis B virus-related insertional mutagenesis in chronic hepatitis B patients as an early drastic genetic change leading to hepatocarcinogenesis. Oncogene, 24: 4340–4348.

Murillo R, Hussey MS, Possee RD. 2011. Evidence for covert baculovirus infections in a Spodoptera exigua laboratory culture. J Gen Virol, 92: 1061–1070.

Perng GC, Chokephaibulkit K, Thompson RL, Sawtell NM, Slanina SM, Ghiasi H, Nesburn AB, Wechsler SL. 1996. The region of the herpes simplex virus type 1 LAT gene that is colinear with the ICP34.5 gene is not involved in spontaneous reactivation. J Virol, 70: 282–291.

Qiu L, Hou L, Zhang B, Liu L, Li B, Deng P, Ma W, Wang X, Fabrick JA, Chen L, Lei C. 2015. Cadherin is involved in the action of Bacillus thuringiensis toxins Cry1Ac and Cry2Aa in the beet armyworm, Spodoptera exigua. J Invertebr Pathol, 127: 47–53.

Saffert RT, Kalejta RF. 2007. Human cytomegalovirus gene expression is silenced by Daxx-mediated intrinsic immune defense in model latent infections established *in vitro*. J Virol, 81: 9109–9120.

Slavokhotova AA, Shelenkov AA, Odintsova TI. 2015. Prediction of Leymus arenarius (L.) antimicrobial peptides based on de novo transcriptome assembly. Plant Mol Biol, 89: 203–214.

Sriram S, Gopinathan KP. 1998. The potential role of a late gene expression factor, lef2, from Bombyx mori nuclear polyhedrosis virus in very late gene transcription and DNA replication. Virology, 251: 108–122.

Sun L, Qiu G, Cui L, Ma C, Yuan H. 2015. Molecular characterization of a ryanodine receptor gene from Spodoptera exigua and its upregulation by chlorantraniliprole. Pestic Biochem Physiol, 123: 56–63.

Tamori A, Yamanishi Y, Kawashima S, Kanehisa M, Enomoto M, Tanaka H, Kubo S, Shiomi S, Nishiguchi S. 2005. Alteration of gene expression in human hepatocellular carcinoma with integrated hepatitis B virus DNA. Clin Cancer Res, 11: 5821–5826.

Tao X, Gu YH, Wang HY, Zheng W, Li X, Zhao CW, Zhang YZ. 2012. Digital gene expression analysis based on integrated de novo transcriptome assembly of sweet potato [Ipomoea batatas (L.) Lam]. PLoS One, 7: e36234.

The UniProt Consortiums. 2011. Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Res, 39: D214–219.

Thompson RL, Sawtell NM. 1997. The herpes simplex virus type 1 latency-associated transcript gene regulates the establishment of latency. J Virol, 71: 5432–5440.

Virto C, Navarro D, Tellez MM, Herrero S, Williams T, Murillo R, Caballero P. 2014. Natural populations of Spodoptera exigua are infected by multiple viruses that are transmitted to their offspring. J Invertebr Pathol, 122: 22–27.

Vogel H, Badapanda C, Knorr E, Vilcinskas A. 2014. RNA-sequencing analysis reveals abundant developmental stage-specific and immunity-related genes in the pollen beetle Meligethes aeneus. Insect Mol Biol, 23: 98–112.

Wang Y, Wu W, Li Z, Yuan M, Feng G, Yu Q, Yang K, Pang Y. 2007. ac18 is not essential for the propagation of Autographa californica multiple nucleopolyhedrovirus. Virology, 367: 71–81.

Weng Q, Yang K, Xiao W, Yuan M, Zhang W, Pang Y. 2009. Establishment of an insect cell clone that harbours a partial baculoviral genome and is resistant to homologous virus infection. J Gen Virol, 90: 2871–2876.

Westenberg M, Uijtdewilligen P, Vlak JM. 2007. Baculovirus envelope fusion proteins F and GP64 exploit distinct receptors to gain entry into cultured insect cells. J Gen Virol, 88: 3302–3306.

Xiao CL, Mai ZB, Lian XL, Zhong J, Jin JJ, He QY, G. Z. 2014. FANSe2: a robust and cost-efficient alignment tool for quantitative next-generation sequencing applications, p. e94250, PLoS One, vol. 9.

Xu K, Sun F, Chai G, Wang Y, Shi L, Liu S, Xi Y. 2015. De novo assembly and transcriptome analysis of two contrary tillering mutants to learn the mechanisms of tillers outgrowth in switchgrass (Panicum virgatum L.). Front Plant Sci, 6: 749.

Xu HJ, Yang ZN, Zhao JF, Tian CH, Ge JQ, Tang XD, Bao YY, Zhang CX. 2008. Bombyx mori nucleopolyhedrovirus ORF56 encodes an occlusion-derived virus protein and is not essential for budded virus production. J Gen Virol, 89: 1212–1219.

Yang Y, Smith. SA. 2013. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. BioMed Central, 14: 328.

Yu M, Carstens EB. 2012. Choristoneura fumiferana multiple nucleopolyhedrovirus LEF-3-P143 complex can complement DNA replication and budded virus in an AcMNPV LEF-3-P143 double knockout bacmid. J Gen Virol, 93: 383–388.

Zhang G, Fedyunin I, Kirchner S, Xiao C, Valleriani A, Ignatova Z. 2012. FANSe: an accurate algorithm for quantitative mapping of large scale sequencing reads. Nucleic Acids Res, 40: e83.

Zhao YJ, Zeng Y, Chen L, Dong Y, Wang W. 2014. Analysis of transcriptomes of three orb-web spider species reveals gene profiles involved in silk and toxin. Insect Sci, 21: 687–698.

Zheng XL, Wang P, Cheng WJ, Wang XP, Lei CL. 2012. Projecting overwintering regions of the beet armyworm, Spodoptera exigua in China using the CLIMEX model. J Insect Sci, 12: 13.