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...
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 genomewide 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 RNaseFree Micro Kit (QIAGEN GmBH, Germany) and an RNaseFree 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 hexamers primers, 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.