Genomic sequencing and analyses of HearMNPV—a new Multinucleocapsid nucleopolyhedrovirus isolated from *Helicoverpa armigera*

Ping Tang1,2, Huan Zhang3, Yinü Li1, Bin Han1, Guozeng Wang1, Qilian Qin3* and Zhifang Zhang1*

**Abstract**

**Background:** HearMNPV, a nucleopolyhedrovirus (NPV), which infects the cotton bollworm, *Helicoverpa armigera*, comprises multiple rod-shaped nucleocapsids in virion (as detected by electron microscopy). HearMNPV shows a different host range compared with *H. armigera* single-nucleocapsid NPV (HearSNPV). To better understand HearMNPV, the HearMNPV genome was sequenced and analyzed.

**Methods:** The morphology of HearMNPV was observed by electron microscope. The qPCR was used to determine the replication kinetics of HearMNPV infectious for *H. armigera* in vivo. A random genomic library of HearMNPV was constructed according to the “partial filling-in” method, the sequence and organization of the HearMNPV genome was analyzed and compared with sequence data from other baculoviruses.

**Results:** Real time qPCR showed that HearMNPV DNA replication included a decreasing phase, latent phase, exponential phase, and a stationary phase during infection of *H. armigera*. The HearMNPV genome consists of 154,196 base pairs, with a G + C content of 40.07%. 162 putative ORFs were detected in the HearMNPV genome, which represented 90.16% of the genome. The remaining 9.84% constitute four homologous regions and other non-coding regions. The gene content and gene arrangement in HearMNPV were most similar to those of *Mamestra configurata* NPV-B (MacoNPV-B), but was different to HearSNPV. Comparison of the genome of HearMNPV and MacoNPV-B suggested that HearMNPV has a deletion of a 5.4-kb fragment containing five ORFs. In addition, HearMNPV orf66, bro genes, and hrs are different to the corresponding parts of the MacoNPV-B genome.

**Conclusions:** HearMNPV can replicate in vivo in *H. armigera* and in vitro, and is a new NPV isolate distinguished from HearSNPV. HearMNPV is most closely related to MacoNPV-B, but has a distinct genomic structure, content, and organization.

**Keywords:** Baculovirus, *Helicoverpa armigera*, Multinucleocapsid nucleopolyhedrovirus, Genome sequence comparison
Background

Members of the family Baculoviridae are rod-shaped viruses with circular, covalently closed, double-stranded DNA genomes [1]. This family includes four genera: *Alphabaculovirus* (lepidopteran-specific nucleopolyhedroviruses (NPVs)), *Betabaculovirus* (lepidopteran-specific granuloviruses), *Gammabaculovirus* (hymenopteran-specific NPVs) and *Deltabaculovirus* (dipteran-specific NPVs) [2]. To date, 54 baculovirus genomes have been sequenced (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10442), including 37 from *Alphabaculovirus*, 13 from *Betabaculovirus*, three from *Gammabaculovirus* and one from *Deltabaculovirus*. Nucleopolyhedrovirus (NPV) and granulovirus (GV) are distinguished from each other by their occlusion body morphology. The NPVs produce large, polyhedron-shaped occlusion bodies, called polyhedra, which contain many virions, whereas the GVs have smaller occlusion bodies, called granules, which normally contain a single virion. The NPVs are further designated as single-nucleocapsid (S) or multinucleocapsid (M), depending on the potential number of nucleocapsids packaged in an envelope of the virion.

The cotton bollworm, *H. armigera*, is a serious pest that causes economic losses to over 60 vegetable and field crops throughout the world [3]. *H. armigera* larvae are significantly resistant to chemical insecticides; therefore, baculovirus pesticides have been recognized as one of the most promising agents to control such pests [4]. HzSNPV was registered as one of the first commercial baculovirus pesticides (Virion-H, Biocontrol-VHZ, Elcar) in the 1970s, and has been extensively used to control the cotton bollworm in the USA [5]. HearSNPV was also the first commercial baculovirus pesticide used to control *H. armigera* in China, and has been extensively used for the control of the pests in vegetable crops [6].

The DNA genomes of HearSNPV-G4, HearSNPV-C1 [8], HearNPV-NNg1 [9], and HzSNPV [10] have been sequenced. Among them, the HearSNPV-G4 and HearSNPV-C1 were isolated from China, HearNPV-NNg1was isolated from Kenya, and HzSNPV was isolated from the United States. Comparative genomic analyses showed that overall gene content and arrangement in these four viruses were highly conserved, and they are considered variants of the same NPV species [9]. In addition, the nucleotide sequence of the HearGV DNA genome was reported [11]. Multinucleocapsid NPVs isolated from *H. armigera* (HearMNPV) producing ODV virions with multiple nucleocapsids per envelope have been identified [12,13]. The genes of other 18 HearMNPV isolates from *H. armigera* which included *lef*-6, *lef*-9, polyhedrin have been reported [14].

In this study, a new nucleopolyhedrovirus isolated from *H.armigera* was observed by electron microscope (EM), suggesting it was multinucleocapsid NPV. Experimental infection of insect larvae indicated that host range of HearMNPV was different from that of HearSNPV and that the cytopathological effect of HearMNPV differed from that of HearSNPV. This report describes the sequence and organization of the HearMNPV genome and compares it with sequence data from other baculoviruses, such as HearSNPV and MacoNPV-B.

Methods

Viruses and insects

HearMNPV was originally isolated from a naturally infected *H. armigera* in the Shanghai city, China in the 1970s. The virus was propagated in laboratory stocks of healthy third instar *H. armigera* larvae by *per os* infection. A laboratory stock of eastern armyworm, cotton leaf worm and beet armyworm were reared at 26°C with a 16:8 h light:dark cycle on a semi-synthetic diet.

Virus purification, DNA extraction, and construction of genomic DNA libraries

To generate a large number of polyhedra, healthy third instar *H. armigera* larvae were inoculated and the hemolymph was collected from the *H. armigera* larvae were collected on ice and centrifuged for 10 min at 4°C. The precipitate was washed several times with distilled water and re-suspended in 0.1% SDS for 30 min at room temperature. After centrifugation, the clean polyhedra were re-suspended in 200 μl TE buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA) [15].

The genomic DNA of HearMNPV was purified according to the following protocol: about 5 × 10⁸ polyhedra were dissolved in 0.1 M Na₂CO₃, 0.15 M NaCl, pH10.4 on ice for 10 minutes, SDS was then added to a final concentration of 0.5%, and the solution was kept on the ice for another 10 minutes. The genomic DNA was extracted twice in an equal volume phenol (pH8.0) and once in chloroform. The DNA was precipitated with two volumes ethanol, washed with 70% ethanol, and dissolved in 0.1 × TE buffer (pH8.0) [16]. The quantity and quality of the isolated DNA were determined by spectrophotometrically and by electrophoresis on 0.7% agarose.

A random genomic library of HearMNPV was constructed according to the “partial filling-in” method and contained 2.0 to 5.0 kbp fragment in vector pUC19 [15,16] DNA fragments for sequencing were prepared from 527 recombinant plasmids. The recombinant plasmids were sequenced with plasmid specific primers and ’primer nesting’ from both strands, using BigDye Terminator v3.1 (ABI) on a 3130XL Genetic analyzer (ABI). The combined sequence was generated from these clones represented a six-fold genomic coverage. The gaps were filled by PCR.
The insect cell lines and infection
The Hz-AM1 cell line and HaBacHZ8-GFP were gifts from Dr. Fei Deng of Wuhan Institute of Virology, Chinese Academy of Sciences. HaBacHZ8 is a bacmid of HearNPV that lacks the polyhedrin gene. An enhanced GFP gene was introduced to HaBacHZ8 by using the HearSNPV bac-to-bac system [17,18] and this generated the bacmid HaBacHZ8–GFP [19]. The QB-Ha-E-5 cell line, which was a gift from Dr. Guiling Zheng of Shandong Agricultural University, was established from the embryonic tissue of H. armigera (Lepidoptera: Noctuidae). The cell line had been subcultured over 60 passages in TNM-FH insect medium supplemented with 10% fetal bovine serum. The cell line could be infected by H. armigera single nucleopolyhedrovirus (HaSNPV) [20]. The Hz-AM1 cells and QB-Ha-E-5 cells were cultured at 27°C in TNM-FH insect medium (Sigma, USA) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (Gibco-BRL, Gaithersburg, USA). Hz-AM1 cells and QB-Ha-E-5 cells were infected with HearNPV at a multiplicity of infection (MOI) of 5. For coinfection, QB-Ha-E-5 cells were infected simultaneously with HearMNVP and HaBacHZ8-GFP at an MOI of 5 for each virus. The cells were examined using Nikon-Ts100 and Leica TCS SP5 II microscopes.

Scanning electron microscopy
Polyhedra were fixed in 2.5% glutaraldehyde at 4°C for 2 h. The fixed sample was dehydrated through a serial series of ethanol solutions and embedded in Spurr’s resin. A diamond knife was used to cut ultrathin sections on a Reichert OME3 Ultramicrotome. The sections were stained with 2% aqueous uranyl acetate, followed by lead citrate. Micrographs of the Polyhedra were taken with a Hitachi S3400N transmission electron microscope at 80 kV.

Transmission electron microscopy
Polyhedra were fixed in 2.5% glutaraldehyde in 0.05 M cacodylate buffer at 4°C for 2 h and post-fixed in 1% osmium tetroxide in the same buffer for 2 h at room temperature. Fixed samples were dehydrated through a graded series of ethanol solutions and embedded in Spurr’s resin. Sections were cut, stained with uranyl acetate and lead citrate, and examined under a JEM-1230 transmission electron microscope (TEM) at an accelerating voltage of 80 kV.

Quantitative PCR (qPCR)
Third-instar larvae werestarved for 12 h at 26°C before being inoculated, and H. armigera test larvae were allowed to ingest a diet soaked in a 10 μl drop, containing an estimated 10^7 OBs. Control larvae ingested a diet soaked in a 10 μl drop with no OBs. The diet soaked OBs or ddH2O were replaced by fresh diet with no OBs after 2 hours adsorption period at 26°C. Time zero of the infection was defined as the time when the diet soaked OBs or ddH2O was removed from the culture boxes. Larvae used in this experiment were sacrificed at various time points ranging from 4 to 96 h post-inoculation (p.i.). A powder was prepared from ten larvae using a mortar and pestle after each collection under liquid nitrogen. 0.1MNa2CO3, 0.15 M NaCl, 1%NP-40 was added to the powders to a total volume of 700 μL. Total DNA was then extracted by the addition of an equal volume of phenol (pH8.0) (twice) and chloroform (once). The DNA was precipitated with two volumes of ethanol, washed with 70% ethanol, and dissolved in ddH2O. The quantity and quality of the isolated DNAs were determined spectrophotometrically.

HearMNVP DNA copy number was determined by real-time qPCR with primers specific to the r2b gene. The viral copy number was then normalized against host-genome copy number by qPCR with primers specific to the host actin gene [21]. The r2b and actin genes were amplified by PCR and cloned into pGEM-T. Recombinant plasmid DNA concentrations were quantified using a spectrophotometer and dilution standards were generated. For each standard dilution, three independent qPCRs were performed using r2b or actin specific primers, and standard curves were generated. For each larval DNA extract, three independent qPCRs were performed using r2b and actin specific primers. The mean of the HearMNVP DNA copy numbers were determined and the number of r2b amplicons was normalized against the number of host actin genes to derive the mean number of viral copies per mean host actin gene copy number. The specific primers were as follows:

actin-F 5’ CTCTTCCAGCCCTCAATTCTTG 3’
actin-R 5’ TTCTGCATACGGTCAGCGATA 3’
r2b-F 5’ AGCAACAAGACCTTAATACTCAACGC 3’
r2b-R 5’ AATATGGCTGCAAAGCTACCG 3’

DNA sequence analysis
Restriction fragments from recombinant plasmids were sequenced and assembled into contigs using SeqMan5.0 from the DNASTAR software package. PCR was used to generate gap-spanning fragments and low quality data regions after preliminary assembly. Open reading frames (ORFs) were identified using ORF Finder http://www.ncbi.nlm.nih.gov/orf/gorf.html [22]. The criterion for defining an ORF was a size of at least 150 nt (50 aa) with minimal overlap. Promoter motifs present upstream of the putative ORFs were screened as described previously [23]. Homology searches were done through the National Centre for Biotechnology Information (NCBI) website using BLAST [24]. Multiple alignments and
percentage identities were performed using the Clustal W. The Tandem Repeats Finder http://tandem.bu.edu/trf/trf.html was used to locate and analyze the homologous regions (hrs) [25]. GeneParityPlot analysis was performed as described previously [26]. A phylogenetic tree was inferred from amino acid sequences by NJ and MP analyses using MEGA, version 5.0 [27]. Bootstrap analyses were performed to evaluate the robustness of the phylogenies using 1000 replicates for both NJ and MP analyses.

Results and discussion

Electron microscopy observation

Scanning electron microscopy revealed that the purified occlusion bodies (OBs) of NPV originating from infected cotton bollworm have irregular shapes, with diameters of about 2 ± 0.3 μm (Figure 1A). Transmission electron microscopy showed multiple rod-shaped nucleocapsids of about 230 nm in length and 50 nm in width embedded in each OB, with multiple nucleocapsids packaged within the envelope of the virion (Figure 1B). These results indicated that the virus is a typical multinucleocapsid NPV. However, transmission electron microscopy indicated that HearSNPV have single nucleocapsids packaged in their virion (Figure 1C). Therefore, the isolate was termed *H. armigera* multinucleocapsid nucleopolyhedrovirus (HearMNPV).

HearMNPV infected insect larvae and cells

Experimental infection of insect larvae showed that HearMNPV can infect the eastern armyworm (*Pseudalestia separate*), but cannot infect either the cotton bollworm (*Spodoptera littura*) or the beet armyworm (*Spodoptera exigua*). By contrast, HearSNPV cannot infect *P. separate*. These results indicate that the host range of HearMNPV differs from that of HearSNPV. Moreover, HearMNPV-infected Hz-AM1 cells produced no polyhedral and showed no typical cytopathic effects (CPE), even at 96 h post-infection (pi) (Figure 2A). However, HearMNPV-infected QB-Ha-E-5 cells produced polyhedra (Figure 2B). It was previously reported that Hz-AM1 cells were permissive to HearSNPV-G4 [8,17,18]. When QB-Ha-E-5 cells were infected by HaBacHZ8-GFP [18,19], which was constructed from HearSNPV, green fluorescence was observed under fluorescence microscopy (Figure 2C). These results indicate that the host range and cells infected by HearMNPV differ from those infected by HearSNPV.

Coinfection of QB-Ha-E-5 cells with HaBacHZ8-GFP and HearMNPV showed that certain cells possessed green fluorescence (under fluorescence microscopy), some cells produced polyhedra, and some cells possessed both green fluorescence and polyhedra (Figure 2C). The results indicate in cells coinfected with two distinct viruses, the viruses are able to coexist, replicate, and package themselves independently. *Cydia pomonella* granulovirus (CpGV) is one of the most successful commercial baculovirus insecticides; however, resistance of the codling moth (*C. pomonella*) to commercially applied CpGV in orchards located in Germany and France has occurred [28]. Therefore, alternating virus treatment or using a mix of HearMNPV and HearSNPV could delay the development of resistance in *H. armigera*, helping to improve both the prevention and control of *H. armigera* in the field.

HearMNPV virus DNA replication in vivo

HearMNPV is a potentially new isolate infectious for *H. armigera* based on the analysis of host range and morphology. Thus qPCR was used to determine the replication efficiency of HearMNPV infectious for *H. armigera* in vivo. Figure 3 shows the quantity of HearMNPV viral DNA in infected larvae, in the decreasing phase (0–4 hr), latent phase (4–12 h), exponential phase (12–48 h), and stationary phase (48–96 h). Initially, the number of viral DNA (vDNA) copies appeared to decrease between 0 h post infection (p.i.) and 4 h p.i. before increasing by 6.97 times between 4 h p.i. and 12 h p.i. The number of vDNA dramatically increased between 12 h p.i. and 48 h p.i., from 452/10⁵ actin to 2.02 × 10¹¹/10⁵ actin, an increase of 4.46 × 10⁸ fold. These results

![Figure 1](https://via.placeholder.com/150)  
**Figure 1** Electron micrographs of polyhedra from HearMNPV and HearSNPV. **A.** Scanning electron micrograph of HearMNPV (12,000x) **B.** Transmission electron micrograph of HearMNPV (80,000x) **C.** Transmission electron micrograph of HearSNPV (20,000x).
indicated that the viral DNA replicated about 29 times, taking about 1.24 h to generate another vDNA copy. This trend continued into the stationary phase, to a lesser degree: vDNA increased 4.82 times between 48 h p.i. and 96 h p.i, and there were about $1.17 \times 10^{12}$ copies per $10^5$ actin at 96 h. These results suggested that *H. armigera* could be infected by HearMNPV efficiently and the replication kinetics conformed what has previously been described for other baculoviruses [29].

**Nucleotide sequence of the HearMNPV genome**

The HearMNPV genome consists of 154,196 bp (GenBank accession no. NC_011615), which is similar to the genomes of MacoNPV-A (155,060 bp) and MacoNPV-B (158,482 bp). The HearMNPV genome has a G+C content of 40.07%, which is within the 58% (LdMNPV) and 32.7% (ChocGV) range for baculovirus genomes, and is similar to MacoNPV-B (40%), AcMNPV (40.7%), BmNPV (40.4%), and EppoNPV (40.7%). According to the adopted convention, the adenine residue at the translation initiation codon of the *polyhedrin* gene represented the zero point on the HearMNPV physical map, and the *polyhedrin* gene was designated as ORF 1 (Table 1). A total of 162 putative ORFs and four homologous regions (*hrs*) were detected in the HearMNPV genome, using computer-assisted analysis to select ORFs starting from a methionine-initiated codon (ATG) and including at least 50 amino acids (aa) and having a minimal overlap with other ORFs [30,31]. All 162 ORFs are shown in Table 1 by location, orientation, size, and potential baculovirus homologs.

HearMNPV ORFs had an average length of 870 bp, with ORF85 (*helicase*) being the largest (3,627 bp) and ORF99 (*ctl*, conotoxin-like protein) being the smallest (150 bp). The 162 predicted ORFs encode 46,677 aa. The total coding sequence and intergenic regions were 139,026 and 15,170 bp, which represented 90.16% and 9.84% of the genome, respectively. The four *hrs* were distributed along the genome, with sizes ranging from 724 to 1,766 bp, and their total sequence was 4,749 bp.
Table 1 List of ORFs in HearMNPV and their Homologous ORFs in the MacoNPV-B, MacoNPV-A, HearSNPV(G4), AcMNPV, AgSeNPV and HearGV

| ORF Name | Position | Prm | length | Homologous ORFs[ORF number, amino acid identity(%)] |
|----------|----------|-----|--------|--------------------------------------------------|
| polh     | 1 > 741  | L   | 246    | MacoNPV-B: 100 1 98.0 1 87.4 8 89.3 1 89.8 1 55.1 |
| polh     | 1629     |     |        | MacoNPV-A: 95 2 73.2 2 26.6 9 27.9 2 33.0 2 16.7 |
| pk1      | 2357     | > 3175 L 522 | 2 99.6 3 87.5 3 48.5 10 38.7 3 58.6 3 34.6 |
| hoar     | 3250     | < 5478 E 742 | 4 97.1 4 74.3 4 20.2 4 27.7 |
|         | 5967     | > 6533 E 188 | 5 96.8 5 67.2 |
|         | adv-e56  | 6626 > 7747 L 373 | 6 100 6 89.3 15 52.6 14 52.6 14 42.1 |
| me53     | 7892     | < 8956 | 354 | 7 99.7 7 86.2 16 30.9 19 24.6 7 47.4 17 22.1 |
| F protein| 9563     | > 11599 L 678 | 8 99.9 9 90.9 133 22.1 8 44.4 26 29.8 |
|         | 11719    | < 12678 E 319 | 9 99.1 9 10.2 9 46.1 |
| gp16     | 12729    | < 13016 L 95 | 10 100 11 97.9 119 30.8 10 65.6 |
| p24      | 13209    | < 13718 L 229 | 11 100 12 92.5 118 50.4 12 22.1 11 25.7 |
|         | 13784    | > 14095 L 103 | 12 100 13 91.3 117a 34.6 12 44.4 |
| lef2     | 14049    | > 14696 | 215 | 13 99.1 14 82.0 117 47.9 6 40.6 13 56.3 33 26.8 |
|         | 14792    | > 15175 E 127 | 14 99.2 15 85.5 21 22.0 |
|         | 15251    | 16435 |        | 16436 < 16732 L 98 | 15 99.0 17 93.4 17 44.7 69 40.5 |
| gp16     | 16798    | > 17397 | 199 | 16 98.5 18 87.9 17 25.7 |
|         | 17550    | > 18260 EL 236 | 117 41.2 19 80.6 18 30.9 |
| chitinase| 18318    | < 20006 L 562 | 19 99.3 22 97.0 21 47.9 21 72.4 21 25.8 13 49.4 21 22.0 |
|         | 20197    | > 21192 | 331 | 20 83 24 77.7 22 47.9 22 52.6 |
|         | 21257    | > 21682 E 141 | 21 99.3 25 86.5 23 22.1 13 47.9 23 22.0 |
|         | 21786    | > 22595 E 269 | 22 98.1 26 78.1 25 22.1 13 47.9 25 22.0 |
|         | 22706    | < 23347 | 213 | 23 95.8 27 91.1 27 25.8 27 49.4 27 22.1 13 47.9 27 22.0 |
|         | 23525    | > 23839 L 104 | 24 99.0 29 64.4 29 22.1 13 47.9 29 22.0 |
|         | 23960    | < 24595 E 211 | 25 97.6 30 69.2 30 22.1 13 47.9 30 22.0 |
|           | hel2     | 25012 < 26379 L 455 | 26 98.9 455 26 98.9 455 26 98.9 |
|         | 26512    | < 28266 | 584 | 27 98.5 32 91.7 61 28.6 105 37.1 20 53.3 62 53.4 |
|         | 28331    | > 29356 | 341 | 28 100 33 97.1 56 46.7 127 56.9 19 81.3 |
|         | 29353    | < 29700 L 115 | 29 100 34 90.6 125 29.0 18 72.0 |
|         | lef1     | 29728 > 30375 | 215 | 30 99.1 35 95.8 124 48.2 14 40.0 17 64.2 80 36.6 |
|         | 30375    | > 31424 L 349 | 31 99.7 36 92.6 123 34.1 13 29.5 16 60.3 102 23.4 |
|         | gp37     | 31476 > 32264 L 262 | 32 99.6 37 95.0 58 62.8 64 58.6 26 72.4 109 43.1 |
|         | 32221    | < 32760 L 179 | 33 98.9 38 92.7 38 45.9 |
|         | egt      | 32828 > 34414 E 528 | 34 99.4 39 94.4 126 50.7 15 48.0 28 73.1 |
|         | 34553    | > 35119 L 178 | 35 98.9 40 94.9 127 22.3 29 53.3 |
|         | 35119    | > 35766 | 215 | 36 98.1 41 90.7 128 28.1 17 30.4 30 45.9 |
|         | 35802    | < 38357 L 851 | 37 99.4 42 88.5 129 26.5 31 45.4 |
|         | 38415    | > 38855 L 146 | 38 98.6 43 64.6 83 31.7 146 43.8 148 34.0 107 46.2 |
|         | 38886    | > 39413 L 175 | 39 99.4 44 82.8 4 26.7 32 36.8 |
|         | pkp      | 39434 > 39943 L 169 | 40 100 45 91.6 130 36.2 24 23.5 33 55.6 |
|         | 39965    | < 40306 L 113 | 41 100 46 96.5 34 52.7 |

Tang et al. Virology Journal 2012, 9:168
http://www.virologyj.com/content/9/1/168
Table 1 List of ORFs in HearMNPV and their Homologous ORFs in the MacoNPV-B, MacoNPV-A, HearSNPV(G4), AcMNPV, AgSeNPV and HearGV (Continued)

| No. | ORF   | HearMNPV | MacoNPV-B | MacoNPV-A | HearSNPV(G4) | AcMNPV | AgSeNPV | HearGV |
|-----|-------|----------|-----------|-----------|--------------|--------|---------|--------|
| 41  | arf1  | 40312    | < 41184   | 290       | 47           | 920    | 131     | 29.2   |
| 42  | pif2  | 40940    | > 42199   | 419       | 43           | 99.3   | 48      | 95.0   |
| 43  | pif1  | 42214    | > 43803   | 529       | 44           | 99.6   | 49      | 93.6   |
| 44  | fgf   | 43800    | > 44045   | 81        | 45           | 100    | 50      | 96.3   |
| 45  | alk-exo | 44080   | < 45174   | 364       | 46           | 99.2   | 51      | 72.9   |
| 46  | 45210  | > 46115  | 301       | 47        | 99.7         | 53     | 89.5    | 40     |
| 47  | 46157  | < 47344  | L         | 395       | 48           | 99.8   | 54      | 90.1   |
| 48  | 47582  | < 47920  | L         | 112       | 49           | 100    | 55      | 93.6   |
| 49  | 47919  | > 49082  | L         | 387       | 50           | 99.7   | 56      | 94.6   |
| 50  | 49121  | > 49522  | L         | 133       | 51           | 99.3   | 57      | 94.7   |
| 51  | 49594  | > 50535  | L         | 313       | 52           | 99.7   | 58      | 94.3   |
| 52  | 50544  | < 51593  | L         | 349       | 53           | 100    | 59      | 73.1   |
| 53  | calyx/pep | 51620  | < 52597   | 325       | 61           | 100    | 60      | 97.5   |
| 54  | 52871  | < 53209  | L         | 112       | 62           | 100    | 62      | 84.1   |
| 55  | 53161  | < 53523  | E         | 120       | 63           | 60     | 63      | 90.8   |
| 56  | 53687  | < 54301  | E,L       | 204       | 64           | 98.5   | 65      | 86.8   |
| 57  | sod    | 54366    | < 54821   | 151       | 65           | 100    | 66      | 96.7   |
| 58  | 54878  | > 55243  | L         | 121       | 66           | 100    | 67      | 86.8   |
| 59  | pif3   | 55269    | > 55880   | 203       | 67           | 98.0   | 68      | 92.6   |
| 60  | 55846  | > 56316  | L         | 156       | 68           | 98.1   | 69      | 87.7   |
| 61  | pagr   | 56365    | > 57819   | 484       | 69           | 99.4   | 70      | 87.4   |
| 62  | 57842  | > 58483  | L         | 213       | 70           | 100    | 71      | 96.8   |
| 63  | 58518  | < 59627  | E         | 369       | 71           | 99.4   | 72      | 95.0   |
| 64  | 59758  | < 61523  | E,L       | 204       | 72           | 93.7   | 73      | 87.3   |
| 65  | dutpase | 62105   | > 62449   | 114       | 73           | 91.2   | 74      | 92.2   |
| 66  | 62531  | > 64309  | E,L       | 592       | 78           | 99.0   | 79      | 96.1   |
| 67  | bro-b  | 64565    | > 65572   | 335       | 74           | 77.5   | 75      | 79.6   |
| 68  | p13    | 65621    | > 66457   | 278       | 75           | 99.6   | 76      | 95.3   |
| 69  | sprT   | 66512    | > 67036   | 174       | 76           | 98.9   | 77      | 89.7   |
| 70  | adv-e66a | 67148   | > 69166   | 672       | 77           | 99.7   | 78      | 97.2   |
| 71  | 69163  | < 69474  | L         | 103       | 78           | 99.0   | 79      | 96.1   |
| 72  | adv-ec43 | 69484   | < 70554   | 356       | 79           | 100    | 80      | 95.2   |
| 73  | 70538  | < 70717  | L         | 59        | 80           | 100    | 81      | 98.3   |
| 74  | vp80   | 70714    | > 72360   | 548       | 81           | 99.3   | 82      | 83.1   |
| 75  | p45    | 72388    | > 73521   | 377       | 82           | 99.7   | 83      | 97.9   |
| 76  | p12    | 73508    | > 73816   | 102       | 83           | 100    | 84      | 89.2   |
| 77  | p40    | 73842    | > 74936   | 364       | 84           | 100    | 85      | 92.6   |
| 78  | p6.9   | 74995    | > 75228   | 77        | 85           | 100    | 86      | 83.3   |
| 79  | lef5   | 75225    | < 76046   | 273       | 86           | 99.6   | 87      | 96.3   |
| 80  | 38 k   | 75945    | > 76847   | 300       | 87           | 99.7   | 88      | 96.0   |
| 81  | vef    | 76886    | > 79432   | 848       | 88           | 99.3   | 89      | 81.8   |
| 82  | bro-c  | 79437    | < 80507   | 356       | 89           | 98.3   | 90      | 84.4   |

Tang et al. Virology Journal 2012, 9:168
http://www.virologyj.com/content/9/1/168
Table 1 List of ORFs in HearMNPV and their Homologous ORFs in the MacoNPV-B, MacoNPV-A, HearSNPV(G4), AcMNPV, AgSeNPV and HearGV (Continued)

| No. | ORF | HearMNPV | MacoNPV-B | MacoNPV-A | HearSNPV(G4) | AcMNPV | AgSeNPV | HearGV |
|-----|-----|----------|-----------|-----------|--------------|--------|---------|--------|
| 83  |     | 80591    | 81019     | 142       | 90           | 99.3   | 91      | 95.8   |
| 84  | odv-e28 | 81052    | 81570     | 172       | 91           | 98.8   | 92      | 97.7   |
| 85  | helicase | 81527    | 85156     | 1209      | 92           | 99.3   | 93      | 96.5   |
| 86  | odv-e25 | 85256    | 85906     | 216       | 93           | 100    | 94      | 94.9   |
| 87  |     | 85903    | 86388     | 161       | 94           | 100    | 95      | 98.8   |
|     |     | 86386    | 87145     | 252       | 95           | 99.6   | 96      | 94.8   |
| 89  |     | 87255    | 87770     | 171       | 96           | 97.7   | 97      | 91.2   |
| 90  | lef4 | 87802    | 89166     | 454       | 97           | 99.6   | 98      | 91.9   |
|     |     | 89165    | 90154     | 329       | 98           | 98.5   | 99      | 79.6   |
| 91  |     | 90256    | 90906     | 216       | 93           | 100    | 94      | 94.9   |
| 92  |     | 91117    | 93555     | 812       | 93           | 100    | 95      | 98.9   |
| 93  |     | 93524    | 94111     | 195       | 95           | 100    | 98      | 91.9   |
| 94  |     | 93936    | 94658     | 240       | 95           | 100    | 99      | 82.8   |
| 95  |     | 94627    | 95628     | 333       | 100          | 100    | 104     | 97.9   |
| 96  |     | 95508    | 95963     | 151       | 100          | 100    | 105     | 92.2   |
| 97  |     | 95965    | 97107     | 380       | 100          | 100    | 106     | 97.4   |
| 98  |     | 97104    | 97256     | 50        | 100          | 100    | 107     | 94.0   |
| 99  |     | 98524    | 99258     | 244       | 100          | 100    | 108     | 97.1   |
| 100 |     | 99307    | 100053    | 248       | 100          | 100    | 109     | 97.1   |
| 101 |     | 100010   | 100825    | 271       | 100          | 100    | 110     | 97.1   |
| 102 |     | 100809   | 101714    | 121       | 100          | 100    | 111     | 97.1   |
| 103 |     | 101173   | 102354    | 393       | 100          | 100    | 112     | 97.1   |
| 104 |     | 11014    | 104675    | 753       | 100          | 100    | 113     | 97.1   |
| 105 |     | 11028    | 107676    | 1000      | 100          | 100    | 114     | 97.1   |
| 106 |     | 107710   | 108099    | 129       | 100          | 100    | 115     | 97.1   |
| 107 |     | 108110   | 108367    | 85        | 100          | 100    | 116     | 97.1   |
| 108 |     | 108459   | 109199    | 246       | 100          | 100    | 117     | 97.1   |
| 109 |     | 109191   | 109736    | 181       | 100          | 100    | 118     | 97.1   |
| 110 |     | 109771   | 110232    | 153       | 100          | 100    | 119     | 97.1   |
| 111 |     | 110287   | 110934    | 215       | 100          | 100    | 120     | 97.1   |
| 112 |     | 110975   | 112033    | 352       | 100          | 100    | 121     | 89.6   |
| 113 |     | 112087   | 112776    | 229       | 100          | 100    | 122     | 89.6   |
| 114 |     | 112859   | 114352    | 497       | 100          | 100    | 123     | 89.6   |
| 115 |     | 114430   | 115017    | 95        | 100          | 100    | 124     | 89.6   |
| 116 |     | 115094   | 117598    | 834       | 100          | 100    | 125     | 89.6   |
| 117 |     | 117622   | 118161    | 179       | 100          | 100    | 126     | 89.6   |
| 118 |     | 118194   | 118469    | 91        | 100          | 100    | 127     | 89.6   |
| 119 |     | 118447   | 118956    | 169       | 100          | 100    | 128     | 89.6   |
| 120 |     | 119494   | 119428    | 159       | 100          | 100    | 129     | 89.6   |
| 121 |     | 119678   | 119947    | 89        | 100          | 100    | 130     | 89.6   |
| 122 |     | 119889   | 120098    | 69        | 100          | 100    | 131     | 89.6   |
| 123 |     | 120224   | 121234    | 336       | 100          | 100    | 132     | 89.6   |

Tang et al. Virology Journal 2012, 9:168
http://www.virologyj.com/content/9/1/168
| ORF   | Location   | Transcriptional Direction | Early Promoters | Late Promoters |
|-------|------------|---------------------------|-----------------|----------------|
| lef10 | 121095    | < 121322                 | L 75            | 3a 50.7        |
|       | 121282    | > 121509                 | L 75            | 50.7 114       |
|       | 121523    | > 122509                 | L 328           | 50.7 114       |
|       | 122514    | < 122987                 | L 157           | 50.7 114       |
|       | 122986    | > 123489                 | L 167           | 50.7 114       |
| hr3   | 123528    | 124601                   |                 | 114 62.7       |
| lop3  | 124860    | > 125717                 | L 285           | 43.3 109       |
|       | 125756    | < 126910                 | L 384           | 43.3 109       |
|       | 126931    | > 127987                 | L 157           | 43.3 109       |
|       | 127986    | > 128489                 | L 167           | 43.3 109       |
| hr4   | 134722    | 135445                   |                 | 134 78.9       |
| iap3  | 124860    | > 125717                 | L 285           | 43.3 109       |
|       | 125756    | < 126910                 | L 384           | 43.3 109       |
|       | 126931    | > 127987                 | L 157           | 43.3 109       |
|       | 127986    | > 128489                 | L 167           | 43.3 109       |
| p47   | 132390    | > 133583                 | L 397           | 43.3 109       |
|       | 135447    | < 136519                 | L 397           | 43.3 109       |
| ubiquitin | 138680  | > 138902                 | L 100           | 43.3 109       |
|       | 138958    | > 139503                 | L 181           | 43.3 109       |
|       | 139858    | > 140210                 | L 100           | 43.3 109       |
|       | 140209    | > 141279                 | L 100           | 43.3 109       |
|       | 141285    | > 141710                 | L 100           | 43.3 109       |
|       | 141751    | > 141196                 | L 100           | 43.3 109       |
| ubiquitin | 143289  | < 145262                 | L 100           | 43.3 109       |
|       | 145343    | > 145594                 | L 100           | 43.3 109       |
|       | 145631    | < 147436                 | L 100           | 43.3 109       |
|       | 147438    | > 148035                 | L 100           | 43.3 109       |
|       | 148114    | > 148392                 | L 100           | 43.3 109       |
|       | 148395    | < 149231                 | L 100           | 43.3 109       |
|       | 149270    | > 149527                 | L 100           | 43.3 109       |
|       | 149529    | < 150914                 | L 100           | 43.3 109       |
|       | 150932    | > 151636                 | L 100           | 43.3 109       |
| rr1   | 151802    | > 154047                 | L 100           | 43.3 109       |

Note. Column 3 indicates ORF location and transcriptional direction on the HearMNPV genome. Column 4 indicates the presence of early (E) and/or late (L) promoters located upstream of the start codon of each ORF. E indicates a TATA sequence followed by a CAGT or CATT mRNA start site sequence 20–40 nucleotides downstream, within 180 bp upstream of the start codon. L indicates the presence of a (A/T/G) TAAG up to 180 bp upstream of the initiation codon.

Table 1 List of ORFs in HearMNPV and their Homologous ORFs in the MacoNPV-B, MacoNPV-A, HearSNPV(G4), AcMNPV, AgSeNPV and HearGV (Continued)
accounting for 3.08% of the genome. Thirty-eight ORFs overlapped with adjacent ORFs by between 1 and 244 bp, with a total of 1485 bp.

Of the 162 ORFs identified in HearMNPV, 21 possessed a consensus early promoter motif (a TATA box followed by a CAGT or CATT motif 20 to 40 bp downstream, and up to 180 bp upstream of the initiation codon). Seventy-one ORFs only contained a late promoter motif ((A/T/G) TAAG up to 180 bp upstream of the initiation codon), and nine had both early and late promoter motifs, which might allow transcription during both the early and late stages of infection. Sixty-one ORFs lacked any recognizable consensus early or late promoter motifs, which might allow transcription during both the early and late stages of infection. Eighty-six ORFs (46%) were oriented in a clockwise direction and 76 ORFs (54%) were in a counter clockwise direction, according to the transcription orientation of the polyhedron gene.

**Comparison of HearMNPV ORFs to other baculoviruses**

The overall gene arrangement and the homology between genes of the HearMNPV and other baculoviruses genomes were compared using Identity-GeneParity analysis [26]. The gene content and organization of HearMNPV were compared with a group I NPV (AcMNPV [32]), Group II NPVs (MacoNPV-B [33] MacoNPV-A [34], HearSNPV-G4 [7] and AgseNPV [35]), and GV (HearGV [11]. HearMNPV shares 117 ORFs with AcMNPV, 161 ORFs with MacoNPV-B, 159 ORFs with MacoNPV-A, 123 with HearSNPV-G4, 147 with AgseNPV, and 89 with HearGV. The average amino acid sequence identities between HearMNPV and AcMNPV, HearMNPV-B, MacoNPV-B, HearSNPV-G4, AgseNPV and HearGV were 44.7%, 98.5%, 90.2%, 41.0%, 58.5%, and 38.6%, respectively.

Comparison of the gene order between HearMNPV and MacoNPV-B revealed a significantly gap between orf 52 to orf 53 in the HearMNPV genome. The gap corresponds to a region of MacoNPV-B comprising orf54, 55, 56, 57, 58, 59, and 60. In addition, the orf66 and orf17 of HearMNPV are homologous to the orf18 and orf117 of MacoNPV-B with 99, 41.2% aa identity, respectively (Table 1). However, the locations of these homologues are not conserved. Relative to each other HearMNPV (x-axis) and MacoNPV-B (y-axis) contain 1 and 8 unique genes, respectively. However, HearMNPV and MacoNPV-B maintain perfect co-linearity in gene content and arrangement (Figure 4A).

The gene arrangement of HearMNPV was also completely collinear with that of MacoNPV-A. The result of the Identity-GeneParity analysis showed that relative to each other HearMNPV (x-axis) and MacoNPV-A (y-axis) contain 2 and 10 unique genes, respectively. There was also high collinearity between HearMNPV and MacoNPV-A (Figure 4B).

In terms of gene content, arrangement, and homology level, HearMNPV is significantly distant from HearSNPV-G4, although they infect the same host, *H. armigera*. Relative to each other HearMNPV (x-axis) and HearSNPV-G4 (y-axis) contain 38 and 18 unique genes, respectively, and these genes are distributed throughout the genomes (Figure 4C). The ‘left’ part of the HearMNPV genome (ORF5-69) displayed a high degree of gene scrambling in the gene parity plot analysis. The homologous ORFs from HearMNPV 70 to 160 are approximately collinear with the HearSNPV-G4 ORFs 8 to 96; however, the direction of the diagonal indicates these regions are inverted, relative to each other, except for HearMNPV ORF102-107 (corresponding to HearSNPV-G4 ORF62-67). (Figure 4C)

Relative to each other HearMNPV (x-axis) and AgseNPV (y-axis) contain 15 and 11 unique genes, respectively. The collinearity between HearMNPV and AgseNPV was higher than that between HearMNPV and HearSNPV-G4, and lower than that between HearMNPV and MacoNPV-B or MacoNPV-A. (Figure 4E)

The collinearity between HearMNPV and AcMNPV from Group I was lower than those between HearMNPV and NPVs from group II (Figure 4D); the parity analysis of HearMNPV’ and HearGV ORFs displayed a much more dispersed pattern (Figure 4F).

**Phylogenetic analysis**

Based on 29 concatenated, conserved genes [36], a phylogenetic tree was estimated for 54 baculoviruses. The results reflected the current systematic assignment of the viruses (Figure 5), indicating that HearMNPV and MacoNPV-B are grouped together and are distinct from HearSNPV-G4 HearSNPV-C1 and HearSNPV-NNg1. In addition, the phylogenetic analysis of three highly conserved genes (lef-8, lef-9, and polh) indicated that the HearMNPV sequences were separated from the other eighteen HearMNPV isolates [14]. These results imply that HearMNPV is a new isolate that differs from HearSNPV.

**Genomic Comparison between HearMNPV and MacoNPV-B**

HearMNPV lacks a 5.4-kb fragment that contains five ORFs

Compared with the MacoNPV-B genome, the HearMNPV genome does not have a 5.4-kb fragment that contains ORF54, 55, 56, 57, and 58 (Figure 6). The nucleotide identities between HearMNPV orf52 1–633 bp, 639–896 bp, and 853–1050 bp and MacoNPV-B orf 53, orf 59, orf 60 are 98%, 98%, and 95%, respectively. Amino acids 147–349 of the protein encoded by HearMNPV orf52 are 100% identical to those of the protein encoded by MacoNPV-B orf 53 and amino acids (aa) 1–65 of the protein encoded by HearMNPV orf52 are 86% identical to the amino acid sequence of MacoNPV-B orf 60 (Figure 6, indicated by the
gray parts in the circles and arrows). However, there was no aa sequence identity between the proteins encoded by HearMNPV orf52 and MacoNPV-B orf 59 (Figure 6, indicated by the black parts in the circles and arrows). The MacoNPV-A genome also lacked the 5.4-kb fragment, suggesting that an insertion in the genome might have lead to the division of ORF59 in MacoNPV-A [33].

According to the sequence analysis of 54 whole genomes of baculoviruses, the 5.4-kb fragment present in MacoNPV-B but not in HearMNPV shared homologous sequences with XecnGV [30] and HearGV [11], by reverse alignment (Table 2). However, this phenomenon was not observed in other genomes. Combined with the phylogenetic analysis (Figure 5), the results suggest that the 5.4-kb fragment was gained during evolution of MacoNPV-B and thus the common ancestor of HearMNPV ORF52 evolved to MacoNPV-B ORF53, ORF59, and ORF60 through gaining 5.4-kb fragment, together with subsequent nucleotide mutations, deletions, and insertions (Figure 6). For recombination to occur, the different viruses species have to be coinfected the same host at the same time. A relatively recent recombination event between ancestors of MacoNPV-B and XecnGV resulted in the insertion of a 5.4-kb fragment from an ancestor of XecnGV into the genome of an ancestor of MacoNPV-B genome, suggesting that these lineages were capable of infecting the same host species at some point during their history [33]. HearMNPV and HearGV could infect the same host cotton bollworm, H. armigera, which provides the opportunity for the natural recombination between two viruses. However, HearMNPV did not gain the 5.4-kb fragment from HearGV by recombination.

HearMNPV ORF66

The nucleotide sequence of ORF66 has high nucleotide sequence similarity to MacoNPV-B's ORF17 and ORF18. Presumably, a mutation gave rise to the division of HearMNPV ORF66 into two open read frames in MacoNPV-B.
Figure 5 Phylogenetic analysis of concatenated amino acid sequence alignments, showing bootstrap values >50% for NJ and MP trees at each node (NJ/MP). The location of HearMNPV is shown in bold. The GenBank accession numbers of each virus are listed after the names.

Figure 6 Comparison of the genome structure of HearMNPV and MacoNPV-B. The left and right arrows represent ORFs in HearMNPV and MacoNPV-B genomes, respectively. The numbers above the arrows represent the names of the ORFs in HearMNPV and MacoNPV-B genomes. The lines between the names of the ORFs represent homologies between the HearMNPV and MacoNPV-B genomes. The black region of arrows in the circle or box represent the nucleotide sequences homologies and the gray region of arrows in the circle or box represent the amino acid sequence homologies. The down arrow indicates the sites where the 5.4-kb fragment is inserted. The letter A indicates that the ORF of HearMNPV has no homolog in the corresponding position of MacoNPV-B. The letter B represents the ORF unique to HearMNPV. Double Vertical Lines represent ORFs that are not in the HearMNPV and MacoNPV-B genomes.
HearMNPV ORF66, located between ORF65 (dutpase) and ORF67 (bro-b), is 1779 bp long and encodes a protein 592aa. The aa sequence identity is 99% between the first 301aa of HearMNPV ORF66’s (874–1779 bp) and MacoNPV-B ORF18 (301aa). However, the genome sequence of HearMNPV ORF66 and MacoNPV-B ORF18 are not collinear (Figure 4A, Figure 6).

Nucleotides 676–968 of HearMNPV ORF66 are 95% identical to MacoNPV-B ORF17; however, the amino acids encoded by this nucleotide sequence did not share amino acid identity with the protein encoded by MacoNPV-B ORF17 because of frameshifts and other mutations of a few nucleotides. There is also no sequence similarity between the N-terminal region (1–675 bp) of HearMNPV ORF66 and MacoNPV-B, either at the nucleotide or amino acid level (shown in the boxes of Figure 6).

The HearMNPV ORF66 protein is 92% identical to the five homologous hr1, hr2, hr3, hr4, and hr5, each of approximately 608aa in size, of Heliothis virescens ascovirus-3e (HvAV-3e) [37], and 85% identical to the proteins encoded by orf34 (564aa) and orf77 (606aa) of Spodoptera frugiperda ascovirus-1a (SfAV-1a) [38]. The comparison showed that these homologous ORFs have four conserved cysteine domains, suggestive of a zinc-binding domain, hypothesized to be a DNA binding domain. This putative domain is found at the C terminus of a large number of transposase proteins, indicated that this might be related to gene duplication in the genome.

Interestingly, we have found an element in the right and left flanking DNA sequences of HearMNPV orf66 had two perfect inverted terminal repeats (ITRs) of 13 nucleotides. Moreover, the tetranucleotide 5′-TTAA-3′, which is very common in transposition of the TTAA family, is duplicated upon this element [39,40] (Figure 7). This indicated that this element could insert exclusively at this insertion site (TTAA). However, the ORF66 has no amino acid identity with piggyBac transposase [41] by blastp analysis. Sequence analysis showed that the right and left flanking DNA sequences of MacoNPVB orf18 lacked the sequence CCTCCTAAGACCC. These results indicated homologous of HearMNPV orf66 in MacoNPVB was split into MacoNPVB orf18 and orf17 during evolution.

Searching for homologs of HearMNPV ORF66 among the baculoviruses revealed that only HearMNPV ORF66 (592aa), MacoNPV-B ORF18 (301aa), HearGV ORF53 (572aa), ORF157 (572aa), ORF157 (576aa), and PsunGV ORF39 (571aa) are homologous ORFs. The phylogenetic analysis indicated that HearGV ORF53, ORF157, and PsunGV ORF39 belong to the same phylogenetic branch, while HearMNPV ORF66, MacoNPV-B ORF18, HvAV-3e hr1–hr5, and SfAV-1a ORF34 and 37 belong to the same phylogenetic clade (Figure 8). HearMNPV and HvAV-3e are both isolated from cotton bollworms, HearMNPV ORF66 and HvAV-3e hr1–hr5 share a flank structure, and have the highest amino acid identity among the homologous genes in baculovirus and ascoviridae to date (excluding unreleased relevant data). This data indicated that these genes might have been exchanged among species and genera.

Table 2 Comparison of ORFs aa identity from 5.4-kb fragment of MacoNPV-B, XecnGV and HearGV

| MacoNPV-B | Homologous (% aa identity) | ORF57 | ORF58 |
|-----------|---------------------------|-------|-------|
| XecnGV    | ORF65(98) | ORF54(98) | ORF55(98) | ORF57(98) | ORF58(98) |
| HearGV    | ORF60(94) | ORF59(94) | ORF57(55) | ORF57(94) | ORF123(89) | ORF54(50) |

Figure 7 Diagram of the region with two DNA sequences flanking a putative Transposase ORF (HearMNPV orf66) of 1 779 bp encoding a protein with 592 amino acids. The tetranucleotide TTAA duplicated is characteristic of a transposition event by a transposable element. ITR represent the inverted terminal repeats.
The genomic differences between HearMNPV and MacoNPV-B are mainly located between hr1 and hr2, including the deletion of the 5.4 kb fragment in HearMNPV and the changes in ORF66, both of which were close to a bro gene (Figure 6).

**HearMNPV ORF17**

The locations of the HearMNPV ORF17 and its homologue in the MacoNPV-B genome are not conserved. HearMNPV ORF17 only has 41.2% aa identity to MacoNPV-B ORF117 (e = 6e-46, with 98% query coverage), while HearMNPV ORF110 was collinear at an aa identity of 97.2% with MacoNPV-B ORF117, indicating that HearMNPV ORF17 has no significant collinearity with the homologous ORF of MacoNPV-B.

**HearMNPV unique ORF**

HearMNPV ORF139 is 264 bp long and encodes a protein of 87 aa. There is an early promoter CATT motif in the 180 bp region upstream of the start codon. Using both BLASTX and BLASTP searching, no homologous protein was found among baculoviruses.

**bro genes**

The occurrence of the baculovirus repeat ORF (bro) gene family is a striking feature in many baculovirus genomes [42]. bro genes are associated with regions of viral genome rearrangement [43]. BmNPV BRO proteins have nucleic acid binding activity that influences host DNA replication and transcription [44]. BRO proteins function as nucleocytoplasmic shuttling proteins that utilize the CRM1-mediated nuclear export pathway [45]. We identified six bro genes dispersed among the genome of HearMNPV and named them bro-a to bro-f, according to the order of their appearance on the linearized genome. There are eight and seven bro genes in MacoNPV-A and MacoNPV-B, respectively. The bro genes are classified into four groups, based on the similarity of the 41-amino acid core domain sequences used for LdMNPV BRO protein classification [46]. HearMNPV bro-c, bro-d, and bro-e belong to group I bro genes, bro-a and bro-b belong to group II bro genes, bro-f belongs to group IV. There is no bro gene corresponding to MacoNPV-B bro-b, which belongs to group III. The HearMNPV genome also lacks homologs of the MacoNPV-A bro-a (group I) and bro-c (group III) genes.

The HearMNPV bro-a, bro-b, bro-c, bro-d, bro-e, and bro-f genes showed aa identities of 83%, 77.5%, 98.3%, 89.6%, 97.4%,
and 98.8% to MacoNPV-B bro-a, -c, -d, -e, -f, -g, respectively. MacoNPV-B bro-b is located in the region of 5.4 kb fragment of MacoNPV-B, which is lack in the HearMNPV genome.

The HearMNPV bro-a gene had an N-terminal region from aa 1 to aa 134 with aa identities of 63% and 95% to MacoNPV-B bro-a and MacoNPV-A bro-b, respectively. The C-terminal region, from aa 135 to aa 331, has aa identities of 98% and 93% to MacoNPV-B bro-a and MacoNPV-A bro-b, respectively. This suggested that bro-a C-terminal regions are the highly conserved portions in these three virus genomes.

HearMNPV bro-f shows high homology to a hypothetical protein P20 [47] from Leucania separata NPV (LeseNPV) and MacoNPV-A bro-h, which both encode 179 aa proteins with amino acid identities of 95% and 98%, respectively. HearMNPV bro-f shows the highest homology to MacoNPV-B bro-g, with an amino acid identity of 98.8%. However, amino acids 1–17 of HearMNPV bro-f are not found in MacoNPV-B bro-g.

When comparing the bro genes of HearMNPV with MacoNPV-B, the lowest aa identity is between HearMNPV bro-b and MacoNPV-B bro-c, at 77.5%. The HearMNPV ORF66 gene is adjacent to HearMNPV bro-f.

Figure 9 Comparison of the hr regions between HearMNPV and MacoNPV-B. Panel A: a deduced consensus sequence of domains A and B from each hr region was used for this alignment. Conserved sequences are indicated with different shading: black indicates 100% conservation, gray >70% conservation, and no shading <70% conservation. Panel B: arrows represent the direction and positions of the repeat-A and repeat-B regions, black boxes represent type A repeats and blank boxes represent type B repeats.
bro-b and has changed much comparing with ORF17 and ORF18, which are closest to MacoNPV-B bro-c.

The differences between HearMNPV bro-c, d, and e and their homologs in MacoNPV-B represent minor nucleotide insertions, deletions, and substitutions.

The bro genes of HearMNPV differed from those of MacoNPV-B in both sequence and number, which indicated that the bro gene region is one of the most important in genomic variation of baculoviruses. The differences between HearMNPV and MacoNPV-B (the 5.4 kb fragment and the location of ORF66) were found in the vicinity of a bro gene. These differences indicated that bro gene might play a role in gene exchange, and, consequently, viral virulence and host range.

hrs
Variable numbers of hr sequences, composed of direct repeats containing a “core” imperfect palindrome and dispersed unevenly among the genome in AT rich intergenic regions, have been identified in most baculovirus genomes [48]. The baculovirus hrs act as enhancers of RNA polymerase II-mediated transcription of baculovirus early promoters [49], as well as functioning as origins of DNA replication in transient replication assays [50,51]. They are also sites of frequent recombinant and rearrangement in baculovirus genomes [52,53]. Four hrs were identified in the HearMNPV genome, with the sizes of 1185 bp (hr1), 1766 bp (hr2), 1074 bp (hr3) and 724 bp (hr4), respectively. The hrs are distributed throughout the HearMNPV genome: between orf14 and orf15, orf63 and orf64, orf130 and orf131, and orf138 and orf139 for hr 1, 2, 3, and 4, respectively. Sequence analysis confirmed that the four hrs comprise two apparent domains with perfect or near-perfect 40 bp palindromes (designated type A) and 31 bp flanking repeats (designated type B) at the head/end of one or both sides of the palindromes (Figure 9A). Each hr repeat sequence comprises two apparent domains (type A and type B) that is similar to that described for MacoNPV-A and MacoNPV-B [33,34]. However, the repeat unit numbers in each hr of HearMNPV was different from MacoNPV-A and MacoNPV-B.

The four hrs of HearMNPV are located at similar positions in the genome as those of MacoNPV-B and MacoNPV-A. Sequence alignment between HearMNPV and MacoNPV-B hrs indicated that these four homologous regions had some insertions/deletions of different sizes, giving rise to identities of 92.0%, 92.3%, 86.6%, 81.9%, respectively. hr1 has three insertions, two of 52 bp and 43 bp that contained only a type A repeat, and one of 147 bp that contained type A and type B repeats. hr2 has two insertions (209 bp and 81 bp) that contained both type A and type B repeats. hr3 has the biggest deletion (489 bp) and another large deletion of 136 bp, which also contained both type A and type B repeats. hr4 has a small deletion of 69 bp that also contains both type A and type B repeats (Figure 9B). The HearMNPV hr4 (724 bp) is shorter than the MacoNPV-B hr4 (1178 bp) occurrence, probably caused by the presence of HearMNPV ORF139, which is adjacent to HearMNPV hr4. HearNPV NNgl contains five hrs (hr1-hr5), similar to HearNPV C1, G4, and HzNPV. The arrangement of these hrs on the genome is almost the same in HearNPV C1, G4 and HzNPV, and it is possible that variability in the hr sequences affect not only progeny virus production, but also the insecticidal activity of the Helicoverpa spp. NPVs [9]. The homologous regions are also suggested to be responsible for the difference in virulence between two Mamestra configurata NPV-A variants, v90/4 and v90/2 [54], indicating that the difference in the organization of the homologous regions of HearMNPV and MacoNPV-B are possibly associated with mechanisms of recombination.

Conclusion
HearMNPV differs significantly from HearSNPV not only in biological properties and morphology, but also in gene content, arrangement, and homology level based on genome sequence comparison, which considered to be different viruses, and not variants of the same virus. Although the average amino acid sequence identity between HearMNPV and MacoNPV-B is 98.5%, but their effective host range are different. Moreover, a 5.4-kb segment of the MacoNPV-B genome which is the apparent result of recombination with an ancestor of XecnGV is absent in the HearMNPV genome, suggesting that the recombination event responsible for the occurrence of this 5.4 kb segment occurred after the divergence of MacoNPV-B and HearMNPV. The location and length of HearMNPV orf66 and MacoNPV-B orf18 are different in their respective genomes. Phylogenetic analysis indicated that these events may occur after MacoNPV-B and MacoNPV-A separated from their ancestor. These distinct differences between HearMNPV and MacoNPV-B may account for their different host range.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
PT carried out the molecular cloning and sequence analysis works and drafted the manuscript. YL, BH and GW participated in the sequence analysis and draft preparation. QQ and ZZ performed experiment design and paper writing. All authors read and approved the final manuscript.

Acknowledgements
This work was supported by grants from the National Natural Sciences Foundation of China (31070139) and the “973” Project (2012CB114602).
References

1. Zanotto PM, Kessing BD, Maruniak JE: Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. J Invertebr Pathol 1993, 62:147–164.
2. Jehle JA, Bilrard GW, Bonning BC, Cory JS, Herriot EA, Roehman GF, Thellmann DA, Thiem SM, Vlak JM: On the classification and nomenclature of baculoviruses: a proposal for revision. Arch Virol 2006, 151:1257–1266.
3. Zalucki MP, Murray DAH, Gregg PC, Fitt GP, Twine PH, Jones C: Ecological of Helicoverpa-Armigera (Hubner) and Hekillothis-Punctigera (Wallengren) in the Inland of Australia - Larval Sampling and Host-Plant Relationships During Winter and Spring. Aust J Zool 1994, 42:329–348.
4. Moscardi F: Assessment of the application of baculoviruses for control of Lepidoptera. Annu Rev Entomol 1999, 44:257–289.
5. Shieh TR: Research, development and application of viral pesticides. Adv Virus Res 1989, 36:315–343.
6. Zhang G: Research, development and application of Helolithis viral pesticide in China. Resour Environ Yangtze Valley 1994, 3:39–44.
7. Chen X, Ukel WF, Tarchini R, Sun X, Sandbirin H, Wang H, Peter S, Zuidema D, Lankhorst RK, Vlak JM, et al: The sequence of the Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus genome. J Gen Virol 2001, 82:241–257.
8. Zhang CX, Ma XC, Guo Z: Comparison of the complete genome sequence between C1 and G4 isolates of the Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus. Virology 2005, 333:190–199.
9. Ogembo GS, Caoli BL, Shikata M, Chayeayonsi S, Kobayashi M, Ikeda M: Comparative genomic sequence analysis of novel Helicoverpa armigera nucleopolyhedrovirus (NPV) isolated from Kenya and three other previously sequenced Helicoverpa spp NPVs. Virus Genes 2009, 39:261–272.
10. Chen X, Zhang W, Wang J, Chen G, Lu A, McCutchen SF, Prential JS, Hermann R, Dolan M, Tingey S, et al: Comparative analysis of the complete genome sequences of Helicoverpa zea and Helicoverpa armigera single-nucleocapsid nucleopolyhedrovirus. J Gen Virol 2002, 83:673–684.
11. Harrison RL, Popham HJ: Genomic sequence analysis of a granulovirus isolated from the Old World bollworm, Helicoverpa armigera. Virus Genes 2008, 36:565–581.
12. Pritchett DW, Scott HA, Young SY: Serological relationships of five nucleopolyhedrosis viruses from lepidopterous species. J Invertebr Pathol 1999, 33:183–188.
13. Gettig RR, McCarthy WJ: Genotypic variation among wild isolates of Helolithis spp nuclear polyhedrosis viruses from different geographical regions. Virology 1982, 117(1):245–259.
14. Rowley DL, Popham HJ, Harrison RL: Genetic variation and virulence of nucleopolyhedroviruses isolated worldwide from the heliothine pests Helicoverpa armigera, Helicoverpa zea, and Helolithis virescens. J Invertebr Pathol 2011, 107(2):112–126.
15. Chen Y, Lin X, Yi Y, Yu Y, Zhang Z: Construction and application of a baculovirus genomic library. Z Naturforchr C 2009, 64:574–580.
16. Nie ZM, Zhang ZF, Wang D, He P, Li F, Song L, Chen Y, Ju L, Yu LL, et al: Complete sequence and organization of Antherea pernyi nucleopolyhedrovirus, a d-rich baculovirus. BMC Genomics 2007, 8:248.
17. Wang HZ, Huang Y, Yi Y, Fang MG, Cheng XW, Vlak JM, Hu ZH: A Novel Expression System of Helicoverpa armigera single-nucleocapsid nucleopolyhedrovirus. Virologica Sinica 2002, 17:319–325.
18. Wang H, Deng F, Pillman GP, Chen X, Sun X, Vlak JM, Hu Z: Cloning of biologically active genomes from a Helicoverpa armigera single-nucleocapsid nucleopolyhedrovirus isolate by using a bacterial artificial chromosome. Virus Res 2003, 97:57–63.
19. Wu D, Deng F, Sun X, Wang H, Yuan L, Vlak JM, Hu Z: Functional analysis of FP25K of Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus. J Gen Virol 2005, 86(9):243924–243944.
20. Zheng GL, Li CY, Zhou HK, Li SW, Li GX, Xue M: Establishment of two new cell lines from the embryonic tissue of Helicoverpa armigera (Lepidoptera:Nocuidae)and their responses to baculovirus infection. Acta Entomol Sinica 2010, 53:167–174.
21. Mangé A, Prudhomme JC: Comparison of Bombyx mori and Helicoverpa armigera cytoplasmic actin genes provides clues to the evolution of actin genes in insects. Mol Biol Evol 1999, 16:165–172.
22. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, Schueler GD, Schriml LM, Sequences E, Tatusova TA, et al: Database resources of the National Center for Biotechnology. Nucleic Acids Res 2003, 31:228–33.
23. Kool M, Vlak JM: The structural and functional organization of the Autographa californica nuclear polyhedrosis virus. Arch Virol 1993, 132:1–16.
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. J Mol Biol 1990, 215:403–410.
25. Lauzon HA, Jamieson PB, Krell PJ, Arif BM: Gene organization and sequencing of the Charistoneura fumiferana defective nucleopolyhedrovirus genome. J Gen Virol 2005, 86:895–961.
26. Hu Z, Arif BM, Jin F, Mattens JW, van Oers MM, Vlak JM: Distinct gene arrangement in the Buzura suppressor single-nucleocapsid nucleopolyhedrovirus genome. J Gen Virol 1998, 79:2841–2851.
27. Tarnura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: MEGAS: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011, 28:2731–2739.
28. Azzer Kaiser S, Frisch E, Undorff-Spahn K, Kienze J, Eberle KE, Gund NA, Reinke A, Zeibitz CP, Hedeg D, Huber J, et al: Rapid emergence of baculovirus resistance in codling moth due to dominant, sex-linked inheritance. Science 2007, 317:1916–1918.
29. Lu HS: Insects virus and virus disease of insect. Beijing: Science Press, 1982.
30. Hayakawa T, Ko R, Okano K, Seong S, Gotto C, Maeda S: Sequence analysis of the Xestia c-nigrum granulovirus genome. Virology 1999, 262:277–297.
31. Ukel WF, van Stienen EA, Heldens JG, Broer R, Zuidema D, Goldbach RW, Vlak JM: Sequence and organization of the Spodoptera exigua multicapsid nucleopolyhedrovirus genome. J Gen Virol 1999, 80:3289–3304.
32. Ayres MD, Howard SC, Kuizo J, Lopez-Ferber M, Posses RD: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus. Virology 1994, 202:586–605.
33. Li L, Donly C, Liu Q, Willis LG, Keddie BA, Erlandson MA, Thellmann DA: Identification and genomic analysis of a second species of nucleopolyhedrovirus isolated from Mamestra configurata. Virology 2002, 297:226–244.
34. Li Q, Donly C, Liu L, Willis LG, Thellmann DA: Sequence and organization of the Mamestra configurata nucleopolyhedrovirus genome. Virology 2002, 294:106–121.
35. Jakubowska AK, Peters SA, Ziemnicka J, Vlak JM, van Oers MM: Genome sequence of an enhancer gene-rich nucleopolyhedrovirus (NPV) from Agrotis segetum: collinearity with Spodoptera exigua multiple NPV. J Gen Virol 2006, 87:537–551.
36. van Oers MM, Vlak JM: Baculovirus genomics. Curr Drug Targets 2007, 8:1051–1068.
37. Agar S, Davis J, Wood D, Wilson P, McGrath A: Sequence and organization of the Helolithis virescens ascovirus genome. J Gen Virol 2007, 88:1120–1132.
38. Bideshi DK, Demattei MV, Rouleau-Bonnin F, Stasiak K, Tan Y, Bigot S, Bigot Y, Fedorici BA: Genomic sequence of Spodoptera frugiperda Ascovirus Tα, an enveloped, double-stranded DNA insect virus that manipulates apoptosis for viral reproduction. J Virol 2006, 80:11791–11805.
39. Cary LC, Goebel M, Corsaro BG, Wang HG, Rosen E, FraserMJ: Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-focus of nuclear polyhedrosis viruses. Virology 1989, 172:156–169.
40. Wang J, Miller ED, Simmons GS, Miller TA, Tabashnik BE, Park Y: piggyBac-like elements in the pink bollworm, Pectinophora gossypiella. Insect Mol Biol 2010, 19:177–184.
41. Carpes MP, Nunes JF, Sampiao TL, Castro ME, Zanotto PM, Ribeiro BM: Molecular analysis of a mutant Anticarsia gemmatalis multiple nucleopolyhedrovirus (AgMNPV) shows an interruption of an inhibitor of apoptosis gene (iap-3) by a new class-II piggyBac-related insect transposon. Insect Mol Biol 2009, 18(6):747–757.
42. Bidesty DK, Renault S, Sresia K, Federici BA, Bigot Y: Phylogenetic analysis and possible function of bro-like genes, a multigene family widespread among large double-stranded DNA viruses of invertebrates and bacteria. J Gen Virol 2003, 84:2531–2544.

43. Willis LG, Seipp R, Stewart TM, Erlandson MA, Theilmann DA: Sequence analysis of the complete genome of Trichoplusia ni single nucleopolyhedrovirus and the identification of a baculoviral photolyase gene. Virusology 2005, 338:209–226.

44. Žemskov EA, Kang W, Maeda S: Evidence for nucleic acid binding ability and nucleosome association of Bombyx mori nucleopolyhedrovirus BRO proteins. J Virol 2000, 74:6784–6789.

45. Kang W, Kunihara M, Matsumoto S: The BRO proteins of Bombyx mori nucleopolyhedrovirus are nucleocytoplasmic shuttling proteins that utilize the CRM1-mediated nuclear export pathway. Virology 2006, 350:184–191.

46. Kuzio J, Pearson MN, Harwood SH, Funk CJ, Evans JT, Slavicek JM, Rohrmann GF: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar. Virology 1999, 253:17–34.

47. Xiao H, Qi Y: Genome sequence of Leucania separata nucleopolyhedrovirus. Virus Genes 2007, 35:845–856.

48. Kool M, Ahrens CH, Erlandson MA, Theilmann DA: Replication of baculovirus DNA. J Gen Virol 1995, 76:2103–2118.

49. Theilmann DA, Stewart S: 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 1992, 187:97–106.

50. Ahrens CH, Pearson MN, Rohrmann GF: Identification and characterization of a second putative origin of DNA replication in a baculovirus of Orgyia pseudotsugata. Virology 1995, 207:572–576.

51. Pearson MN, Rohrmann GF: Lymantria dispar nucleopolyhedrosis virus homologous regions: characterization of their ability to function as replication origins. J Virol 1995, 69:219–221.

52. Hayakawa T, Rohrmann GF, Hashimoto Y: Patterns of genome organization and content in lepidopteran baculoviruses. Virology 2000, 278:1–12.

53. Harrison RL, Bonning BC: Comparative analysis of the genomes of Rachipusia ou and Autographa californica multiple nucleopolyhedroviruses. J Gen Virol 2003, 84:1827–1842.

54. Li L, Li Q, Willis LG, Erlandson M, Theilmann DA, Donly C: Complete comparative genomic analysis of two field isolates of Mamestra configurata nucleopolyhedrovirus-A. J Gen Virol 2005, 86:91–105.

doi:10.1186/1743-422X-9-168
Cite this article as: Tang et al.: Genomic sequencing and analyses of HearMNPV—a new Multinucleocapsid nucleopolyhedrovirus isolated from Helicoverpa armigera. Virology Journal 2012 9:168.