


Article

Genome Analysis of a Novel Clade b *Betabaculovirus* Isolated from the Legume Pest *Matsumuraeses phaseoli* (Lepidoptera: Tortricidae)

Ruihao Shu ¹, Qian Meng ¹, Lin Miao ¹, Hongbin Liang ², Jun Chen ², Yuan Xu ²,
Luqiang Cheng ^{1,3}, Wenyi Jin ^{1,4}, Qilian Qin ¹ and Huan Zhang ^{1,*} 

¹ State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; shuruihao@ioz.ac.cn (R.S.); mengqian@ioz.ac.cn (Q.M.); miaolin@ioz.ac.cn (L.M.); chenglq1995@gmail.com (L.C.); jwy707508523@126.com (W.J.); qinql@ioz.ac.cn (Q.Q.)

² National Animal Collection Resource Center, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; lianghb@ioz.ac.cn (H.L.); chenjj@ioz.ac.cn (J.C.); brjdgryx@163.com (Y.X.)

³ College of Horticulture and Plant Protection, Yangzhou University, Yangzhou 225009, China

⁴ College of Life Sciences, Hebei University, Baoding 071002, China

* Correspondence: zhanghuan@ioz.ac.cn

Received: 25 August 2020; Accepted: 20 September 2020; Published: 23 September 2020



Abstract: *Matsumuraeses phaseoli* is a Lepidopteran pest that primarily feeds on numerous species of cultivated legumes, such as *Glycine* and *Phaseolus*. It is widely distributed in northeast Asia. A novel granulovirus, designated as *Matsumuraeses phaseoli granulovirus* (MaphGV), was isolated from pathogenic *M. phaseoli* larvae that dwell in rolled leaves of *Astragalus membranaceus*, a Chinese medicinal herb. In this study, using next-generation sequencing, we report the complete genome of MaphGV. MaphGV genome comprises a double-stranded DNA of 116,875 bp, with 37.18% GC content. It has 128 hypothetical open reading frames (ORFs). Among them, 38 are baculovirus core genes, 18 are lepidopteran baculovirus conserved genes, and 5 are unique to *Baculoviridae*. MaphGV has one baculovirus repeat ORF (*bro*) and three inhibitors of apoptosis proteins (*iap*), including a newfound *iap-6*. We found two atypical baculoviral homologous regions (*hrs*) and four direct repeats (*drs*) in the MaphGV genome. Based on phylogenetic analysis, MaphGV belongs to Clade b of *Betabaculovirus* and is closely related to *Cydia pomonella granulovirus* (CpGV) and *Cryptophlebia leucotreta granulovirus* (CrleGV). This novel baculovirus discovery and sequencing are invaluable in understanding the evolution of baculovirus and MaphGV may be a potential biocontrol agent against the bean ravaging pest.

Keywords: granulovirus; baculovirus genome; biological control; pest management

1. Introduction

Baculoviruses are double-strand DNA viruses that specifically infect the larvae of insect orders such as Lepidoptera, Hymenoptera, and Diptera [1]. The family *Baculoviridae* is divided into four genera. There is much less genomic information for *Beta-*, *Gamma-*, and *Deltabaculovirus* than for *Alphabaculovirus*. The occlusion bodies (OBs) of *Betabaculovirus* are ovocylindrical structures, generally $0.12 \times 0.50 \mu\text{m}$ in diameter, and are significantly smaller than those of *Alphabaculovirus* (0.15 to $5 \mu\text{m}$) [2]. The genus was initially called granuloviruses (GVs) due to the granule-like morphology of the OBs [3]. Each virion of occlusion-derived virus (ODV) typically contains a single nucleocapsid within a single envelope. Baculoviruses typically have narrow host ranges, often limited to just one or a few related insect species [4] and have ancient coevolutionary interactions with their hosts [5]. GVs are confined to the

Lepidoptera and are reported to infect 148 hosts in this family [6], much less than members of the Lepidoptera family [7]. This suggests that numerous betabaculoviruses are yet to be discovered.

Matsumuraeses phaseoli (Lepidoptera: Tortricidae) is a lepidopteran pest that feeds on cultivated legumes, such as *Glycine* and *Phaseolus*. It characteristically curls up and feeds on the leaves of the plant [8]. The host insect is widespread across northeast Asian countries, such as Japan, North Korea, Russia, and China [9]. In China, the species is an important pest for 11 species of economically important leguminous crops, including vetch, broad bean, alfalfa, and clover [10]. In this study, eight *M. phaseoli* larvae presenting symptoms of baculovirus infection were collected from the glued leaves of milkvetch, *Astragalus membranaceus*. *A. membranaceus* is a perennial legume widely cultivated in China; its roots are exploited for traditional Chinese herbal medicine. In recent years, protracted cultivation has resulted in the unprecedented emergence and occurrence of pests. The use of selective pesticides, such as baculovirus-based products, is particularly important to avoid pesticide residues in food crops and plant-based traditional medicines. Baculoviruses infect and kill insect pests in the field. However, most baculoviruses are host species-specific; therefore, novel discovery and sequencing of viruses specific to insect pests are important. Although granuloviruses pathogenic to *M. phaseoli* have been reported in the past decades [11,12], no research has been documented.

In this study, several *M. phaseoli* larvae presenting symptoms of baculovirus infection were collected from glued leaves of *A. membranaceus*. After characterization, we observed the complete genome of a novel granulovirus isolated from *M. phaseoli* larva with symptoms of baculovirus infection. The host species was identified using morphological characteristics and mitochondrial cytochrome c oxidase I (COI) barcode sequences. Therefore, the putative granulovirus was designated as *Matsumuraeses phaseoli granulovirus* (MaphGV).

2. Materials and Methods

2.1. Virus Collection and Host Identification

Eight host insect cadavers displaying symptoms of baculovirus infection were directly collected from *A. membranaceus* leaves in northwest China (Min County, Dingxi city, Gansu province) in September 2019. The bodies of the host insects containing the virus were directly extracted using the phenol-chloroform method. The target mitochondrial COI barcode was amplified by polymerase chain reaction (PCR) using lepidopteran COI primers (Table S1) [13]. The amplified fragments were Sanger sequenced and analyzed by BLASTn [14] with the barcode sequence in BOLD [15] and GenBank [16].

2.2. Morphological Characterization of the Occlusion Bodies

Typical granulovirus features were observed under the light microscope (Leica DM2000, Leica, Wetzlar, Germany). Pure OBs were obtained after several differential centrifugations. A suspension of the OBs was prepared [17] and detected using a scanning electron microscopy (SEM) (Hitachi SU8010, Hitachi, Tokyo, Japan) at a 5 kV acceleration voltage [18]. The size of the MaphGV OB was measured based on 44 complete image samples captured by SEM.

2.3. Genomic DNA Sequencing and Sequence Analysis

Viral genomic DNA was extracted after cleavage of purified OBs in an alkaline solution [19]. The DNA was sequenced using the Illumina HiSeq X system. The high-quality pair-end reads were assembled de novo into contigs using the Iterative Virus Assembler (IVA, version 1.8) [20]. To validate two regions with low mapping rate, PCR amplification, molecular cloning, and subsequent Sanger sequencing were performed. The primers used are listed in Table S1. The strategy allowed for confirmation and correction of the sequences, including the addition of two nucleotides and the elimination of the other six [19]. The genetic variability of the viral isolate was investigated using a bacsnp pipeline [21].

FGENESV [22] and VIGOR [23] were used to predict the hypothetical open reading frames (ORFs) of the MaphGV genome, with at least 50 codons and minimal overlap. The complete genome and annotation information for MaphGV were submitted to GenBank under accession number MT844067. EMBOSS stretcher [24] was used to compare the MaphGV genome with that of related species. Gene parity plots were drawn to assess the pairwise ORF synteny between MaphGV and the selected baculoviruses. [25].

2.4. Non-Coding Region Analysis

Tandem Repeats Finder [26], REPuter [27], Blast2seq [28], and EMBOSS palindrome [29] were used to search the homologous repeat regions (*hrs*) and direct repeats (*drs*) in the MaphGV genome. The secondary structures of repeat sequences were predicted using the ViennaRNA secondary structure server [30]. The structures of *hrs* were visualized by the IBS online server [31]. The alignment of *hrs* repeat units was visualized using BOXSHADE [32].

To screen early and late promoter motifs, we checked the 180 nt upstream region of each initiation codon. Two potential TATA-box elements were derived by SeqKit [33] for early promoters. One had a common TATA-box motif (TATAW) with a CAKT mRNA start site sequence of 25–35 nt downstream [34,35]. The other was derived from TATA-like elements (TAATWAA) initially found upstream of some *lef* ORFs [36]. DTAAG was treated as a baculovirus late promoter element located, on average, –60 nt upstream of the ORF initiation codon [37].

2.5. Phylogenetic Analysis

Protein sequences of 38 baculovirus core genes were extracted from 108 sequenced baculovirus genomes (including MaphGV, Table S2). The sequences were aligned by MAFFT [38] with default parameters. Poorly aligned sites were removed using BMGE [39]. The aligned sequences were then concatenated in the same order as those in the *Autographa californica multiple nucleopolyhedrovirus* (AcMNPV) genome. The phylogenetic relationship was constructed using MEGA version X, based on the minimum evolution method [40].

Kimura two-parameter (K2P) distances [41] were calculated separately for the *granulin*, *lef-8*, and *lef-9* alignments using the nucleotide distance calculation function of R package “ape” (version 5.3) [42]. Substitution rates among sites were set to be uniform. Gaps within the alignment were treated as pairwise deletions.

IQ-TREE (Version 1.6.1) [43] was used to construct the phylogenetic tree for the inhibitor of apoptosis protein (*iap*), based on the maximum likelihood method. The selected substitution model was LG + F + R10.

3. Results and Discussion

3.1. The Host Determination and Virus Characterization

The host insects were classified in *Matsumuraeses* genus (Lepidoptera: Tortricidae) according to morphology and damage symptoms of the leaf rolling of the larva. The symptomatic larva presenting pathological features of granulovirus infection is shown in Figure 1A. The sequenced mitochondrial COI barcoding (GenBank accession number: MT578848.1) confirmed the host to be *M. phaseoli*. Based on historical tradition, baculoviruses were named according to the host species from which they were first isolated, followed by their OB morphology group [44]. Accordingly, the virus was named *Matsumuraeses phaseoli granulovirus* (MaphGV).



Figure 1. Host larva displaying symptoms of MaphGV infection and the corresponding occlusion bodies. (A) *Matsumuraesia phaseoli* larva in the glued *Astragalus membranaceus* leaves exhibiting symptoms for baculovirus infection. (B) MaphGV occlusion bodies as seen under scanning electron micrograph. The scale of measurement is marked in the bottom right corner.

Ultrastructural analysis revealed that MaphGV OBs are ovocylindrical in shape, a characteristic typical of betabaculovirus occlusion bodies (Figure 1B). They were approximately $(395 \pm 20) \times (239 \pm 24)$ nm, which is within the range of dimensions previously reported for *Betabaculovirus* OBs [44].

3.2. Genome Feature

Using the Illumina HiSeq X sequencing system, 4,868,861 high-quality pair-end reads of MaphGV samples were generated. The complete genome of MaphGV was assembled using IVA (version 1.8). The ambiguous regions were further validated by PCR and Sanger sequencing. Only 11 potential single nucleotide polymorphism (SNP) sites were identified, indicating high purity and low genetic diversity of the isolate (Figure S1). The final MaphGV genome consists of 116,875 bp with 128 hypothetical ORFs (Figure 2). The coding regions cover 89.61% of the genome. The gene encoding granulin was designed as the first ORF, with its start codon designed as the first three nucleotides of the genome. MaphGV contains 38 core genes conserved in all baculoviruses, 16 lepidopteran baculovirus conserved genes [45], and 5 hypothetical unique genes without homologs in *Baculoviridae* (Table 1).

Table 1. Gene contents of MaphGV.

Gene Type	Core Genes	Lepidoptera Baculovirus Conserved Genes	Other Baculovirus Genes
Structure	<i>odv-e18</i> (orf13), <i>49k</i> (orf14), <i>odv-ec43</i> (orf47), <i>p48/p45</i> (orf74), <i>p40/bv/odv-c42</i> (orf76), <i>p6.9</i> (orf77), <i>38k</i> (orf79), <i>odv-e25</i> (orf82), <i>p18</i> (orf83), <i>p33</i> (orf84), <i>vp39</i> (orf87), <i>odv-ec27</i> (orf88), <i>ac81</i> (orf93), <i>gp41</i> (orf94), <i>ac78</i> (orf95), <i>desmop</i> (orf100), <i>ac53</i> (orf120), <i>vp1054</i> (orf123)	<i>granulin</i> (orf1), <i>pk-1</i> (orf3), <i>calyx/pep-1</i> (orf20), <i>F protein</i> (orf27), <i>bv-e31</i> (orf61), <i>p24</i> (orf62), <i>p12</i> (orf75), <i>tlp-20</i> (orf92), <i>fp25k</i> (orf118)	<i>p10</i> (orf17), <i>pep-2</i> (orf23), <i>vp80</i> (orf24)
Replication	<i>lef-2</i> (orf34), <i>lef-1</i> (orf64), <i>helicase</i> (orf81), <i>dna-pol</i> (orf99), <i>alk-exo</i> (orf114)	<i>lef-11</i> (orf51), <i>dbp</i> (orf71), <i>lef-3</i> (orf102), <i>me53</i> (orf128)	<i>ie-1</i> (orf6), <i>dna ligase</i> (orf108), <i>helicase-2</i> (orf115)
Transcription	<i>p47</i> (orf60), <i>lef-5</i> (orf78), <i>lef-4</i> (orf86), <i>vlf1</i> (orf96), <i>lef-9</i> (orf106), <i>lef-8</i> (orf117)	<i>39k</i> (orf50), <i>lef-6</i> (orf70)	
Oral infection	<i>pif-5</i> (orf16), <i>pif-3</i> (orf30), <i>pif-2</i> (orf41), <i>ac110</i> (orf45), <i>p74</i> (orf53), <i>pif-1</i> (orf65), <i>pif-4</i> (orf80), <i>pif-8</i> (orf91), <i>pif-6</i> (orf103)		<i>odv-e66</i> (orf31)
Auxiliary gene		<i>38.7k</i> (orf63)	<i>chitinase</i> (orf9), <i>iap-3</i> (orf11), <i>gp37</i> (orf12), <i>bro</i> (orf26), <i>p13</i> (orf39), <i>ubiquitin</i> (orf46), <i>sod</i> (orf52), <i>fgf-1</i> (orf66), <i>iap-1</i> (orf85), <i>iap-2</i> (orf105), <i>fgf-2</i> (orf111), <i>fgf-3</i> (orf126), <i>egt</i> (orf127)
Unique gene		<i>orf22</i> , <i>orf40</i> , <i>orf56</i> , <i>orf101</i> , <i>orf124</i>	

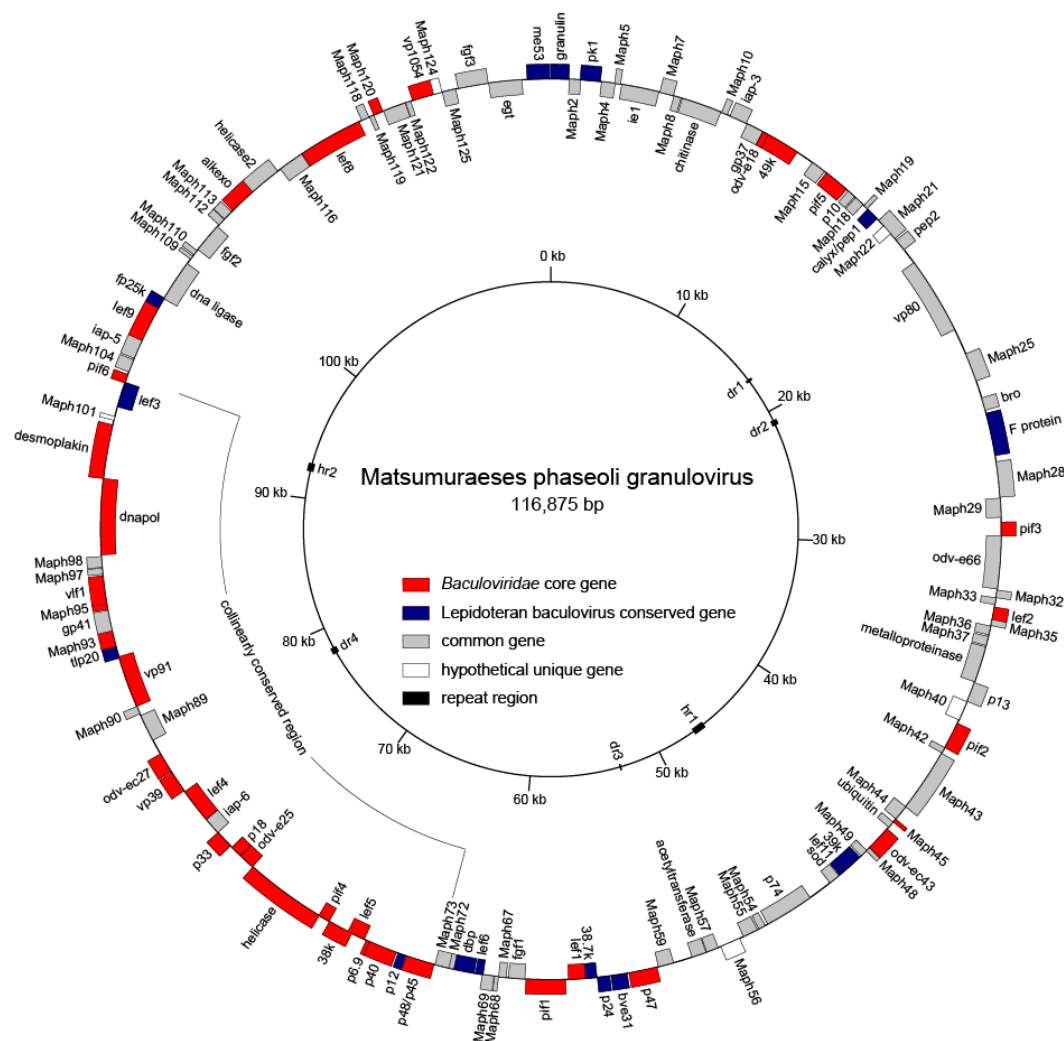


Figure 2. Whole-genome map for MaphGV. Genes in the positive strand are drawn to the outside of the circle and those in the negative strand are on the inside. The gene types are coded as follows: red for core genes, blue for lepidopteran conserved genes, white for hypothetical unique genes, and grey for other common baculovirus genes. Black squares in the inner circle represent repeat regions. The arc indicates the collinearly conserved region.

Among the 128 hypothetical ORFs, 123 have homologs in other baculoviruses, including 30 related to virus structure and assembly, 12 to DNA replication, 8 to transcription, 10 to oral infection, and 14 auxiliary genes (Table 1). MaphGV shares 112 homologous ORFs with *Pieris rapae granulovirus* (PrGV), 109 with *Cydia pomonella granulovirus* (CpGV), 105 with *Cryptophlebia leucotreta granulovirus* (CrleGV), 104 with *Choristoneura occidentalis granulovirus* (ChocGV), 95 with DisaGV, 56 with AcMNPV, 35 with NeseNPV, and 17 with CuniNPV. No *cathepsin*, *enhancin*, or *p35* genes were found.

As for the five unique genes, no homolog in GenBank was found through BLASTp search for the following: *orf22* (144 aa), *orf40* (276 aa), *orf56* (280 aa), *orf101* (55 aa), and *orf124* (117 aa). We screened the regions 180 nt upstream of these putative ORFs for promoter elements. The CAKT motif is an initiator element that was found approximately 30 nt downstream away from the TATA-box (RNA-polymerase-II-binding site) at the transcription start site of many genes [34]. *Orf40* harbors a conserved early promoter pattern (TATA-box with the CAKT motif 25-nt downstream), a *lef*-TATA-like early promoter motif (TAATWAA) [36] and a late promoter (DTAAG) [37]. Both *orf22* and *orf56* possess late promoters [37]. The former harbors a late promoter proximal to the initiation codon (within 15

nucleotides). *Orf101* and *orf124* did not match the elements mentioned above. Further studies are required to explore whether these are functional in MaphGV.

3.3. Phylogenetic Analysis of MaphGV

A phylogenetic tree was constructed based on 38 concatenated predicted amino acid sequences of core genes from 108 completely sequenced baculoviral genomes (including MaphGV). Based on the phylogenetic tree, MaphGV was grouped in clade b of *betabaculovirus* (Figure 3). It is a sister species related to the cluster formed by CpGV and CrleGV, which are isolated from Tortricidae, with 62% bootstrap support. Identification into baculovirus species was based on pairwise nucleotide distances estimated using the K2P model of nucleotide substitution for *polyhedrin/granulin*, *lef-8*, and *lef-9* genes [46,47]. The K2P pairwise distances of the above genes for MaphGV to other granuloviruses were more than 0.05 substitutions/site (Table S3), effectively validating the classification of MaphGV as a novel betabaculovirus species.

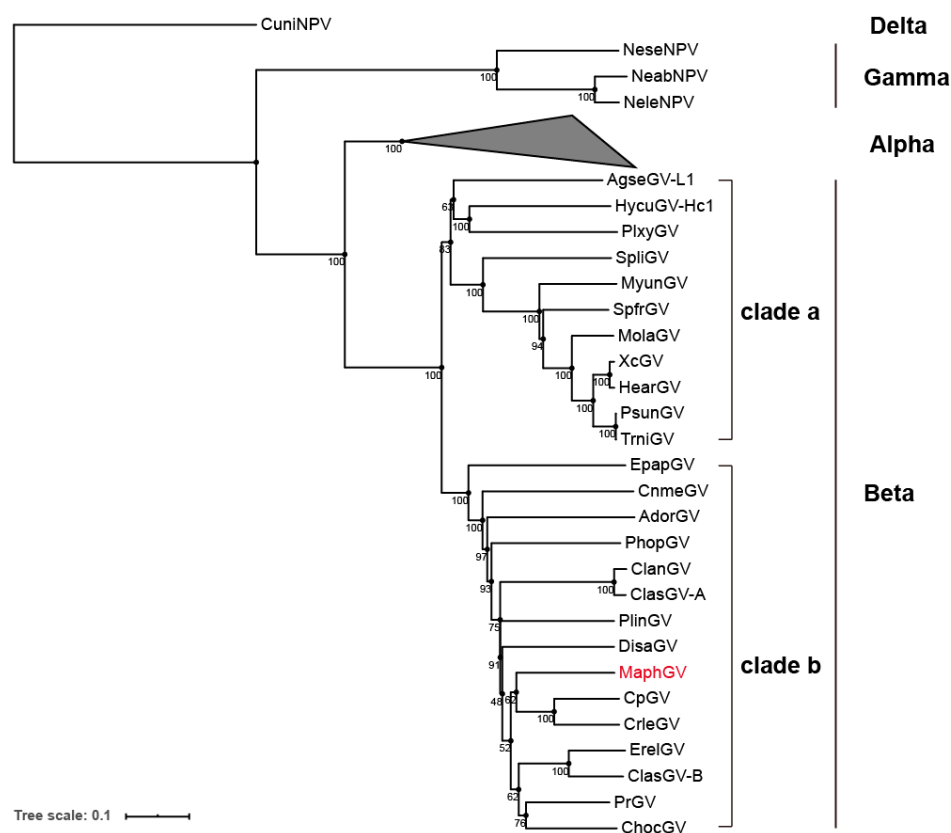


Figure 3. Phylogenetic tree inferred from concatenated predicted amino acid sequences of 38 conserved core genes of 108 sequenced baculovirus genomes. The unrooted tree was generated using the minimum evolution method with 1000 bootstrap values. MaphGV is highlighted in red. Branches of the genera *Alphabaculovirus* are collapsed. Taxon genera are indicated behind corresponding branches. Both clades a and b of *Betabaculovirus* are indicated in brackets.

Gene content mapping revealed the patterns of baculovirus evolution and highlighted the fluid nature of baculovirus genomes [48]. The gene order of the MaphGV genome was compared with other related betabaculoviruses and representative baculoviruses, including AcMNPV (*Alphabaculovirus*, α), *Neodiprion sertifer nucleopolyhedrovirus* (NeseNPV, *Gammabaculovirus*, γ), and *Culex nigripalpus nucleopolyhedrovirus* (CuniNPV, *Deltabaculovirus*, δ) using the gene parity plot. Gene parity plots allow for comparison of gene organization between two different genomes and are used to illustrate collinearity among baculovirus genomes [25]. The gene order of MaphGV was compared with those

of CpGV, CrleGV, PrGV, ChocGV, DisaGV, AcMNPV, NeseNPV, and CuniNPV. The whole-genome nucleotide identity of the MaphGV genome to that of CpGV, CrleGV, PrGV, and DisaGV was 53.8%, 54.8%, 56.8%, and 49.5%, respectively. However, MaphGV shares highly collinear gene order with CrleGV, CpGV, PrGV, and ChocGV, and partial collinearity with AcMNPV (α) (Figure 4). In contrast, its gene arrangement is significantly different from that of NeseNPV (γ) and CuniNPV (δ) (Figure 4). Consistent with previous studies [49], *Betabaculovirus* shares a conserved collinear gene arrangement. Compared with AcMNPV, the genome region between *p48* and *pif-6* is still collinearly conserved (Figure 4).

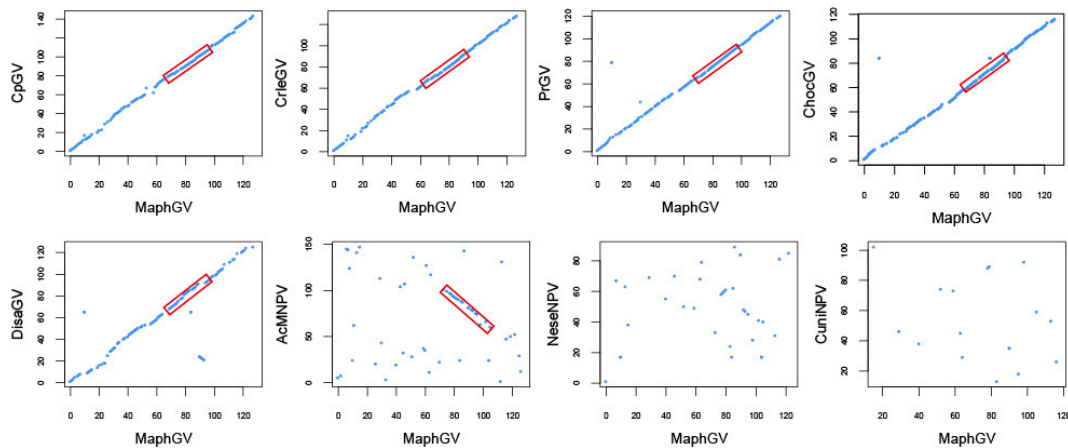


Figure 4. Gene parity plots comparing MaphGV with other baculoviruses. The ORF content and order of the MaphGV genome (x-axis) was compared with that of CpGV, CrleGV, PrGV, ChocGV, DisaGV, AcMNPV (*Alphabaculovirus*), NeseNPV (*Gammabaculovirus*), and CuniNPV (*Deltabaculovirus*) (y-axes). Each point in the plot represents a MaphGV ORF homolog found in other baculoviruses. The red boxes indicate the lepidopteran baculovirus collinearly conserved regions.

3.4. Repeat Sequences

hrs may participate in replication origin of NPVs and GVs and function as enhancers of RNA-polymerase-II-mediated transcription of baculovirus early promoters in NPVs. An individual NPV *hr* typically comprises of a 60–80 bp repeat unit, centered around a palindrome. Unlike NPVs, *hrs* in GVs are more variable and often lack palindromes [49]. For *Betabaculovirus*, conserved *hrs* were found in CpGV, CrleGV, AdorGV, ChocGV, Tortricidae hosts, and PhopGV (Gelechiidae host) with 13 bp imperfect palindrome ends [49]. Although MaphGV is related to CpGV and CrleGV, no typical *hrs* were found.

As a betabaculovirus for tortricids, MaphGV contains two *hrs*, but their structures are atypical (Figure 5A). *hr1* was found in the intergenic region between *sod* and *p74*, with five highly conserved repeat units. *hr2* is located between desmoplakin and *lef-3* and is composed of three repeat units. There is a hypothetical ORF (*orf101*) located in *hr2*. However, no homolog in *Baculoviridae* was found for *orf101*. The repeat units of *hr* are highly conserved within MaphGV, and much longer than typical GV *hr* repeats (104 bp vs. 56–65 bp, Figure 5B). The BLASTn search for the *hr* repeat unit did not reveal any significant hits within an E-value <0.01. In MyunGV [50] and XecnGV- α 4 [51], some of the *hr* units are directly repeated, and the other repeat units occur in the reverse orientation, thus forming an imperfect palindrome. Both *hrs* in MaphGV harbor a 12-bp palindromic core GTAAACGTTTAC between the opposite direction repeat units. They likely occur in equilibrium between the double-stranded DNA and opposite hairpin-loops (*hr1*, −504.14 kcal/mol and *hr2*, −201.18 kcal/mol, Figure 5C) constituted by each complementary strand, thus forming a cruciform-like structure. *hr1* and *hr2* are conserved in the MaphGV genome based on positions relative to conserved ORFs [52].

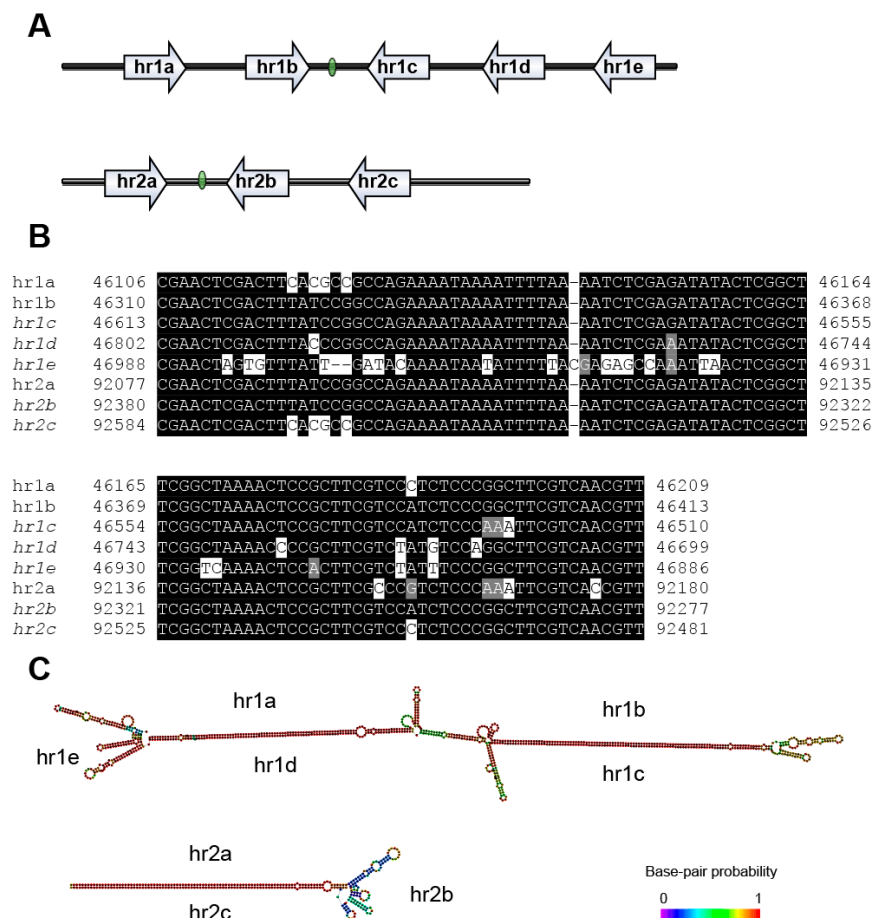


Figure 5. The structure and alignment of MaphGV homologous repeat regions (*hrs*). (A) The schematic diagram for the two *hrs* in the MaphGV genome. The blank arrows indicate the location and direction of the *hr* repeat units and the green bead-shaped region indicates the palindromic core. (B) Alignment of MaphGV *hr* repeat units. Nucleotide positions of the repeat units in the genome sequence are indicated at both ends of the aligned sequences. The name of the *hr* repeat units occurring on the reverse strand (*hr1c–e*, *hr2b–c*) is indicated in italics beside the alignment. The residues identical to the column-consensus (>50% identical) are shaded in black, whereas gray indicates nucleotides of the same class (containing either a purine or pyrimidine base). (C) Predicted secondary structures of MaphGV *hrs*. Secondary structures of single-strand *hrs* DNA were predicted using the ViennaRNA server. The base-pair probabilities of the predicted secondary structure are shown in color.

Intergenic repeat sequence analysis also identified four *drs*. The four *drs* were AT-rich, located in distinct intergenic regions (*pep-2-vp80*, *vp80-orf25*, *orf58-orf59*, *orf88-orf89*) with different repeat units (Table S4).

One to 16 copies of the baculovirus repeat ORF (*bro*) are present in some betabaculoviruses and all sequenced alpha- and deltabaculoviruses [53]. *bro* comprises a highly repetitive and conserved gene family whose function is unclear [53]. In this study, only one *bro* gene (*orf26*) was found in the MaphGV genome.

3.5. *iap-6* Found in MaphGV Genome

MaphGV contains three *iap* lineages. The *iap* gene is recognized as a central player in regulating apoptosis and many other important processes [54]. IAPs always display at least one baculoviral IAP repeat (BIR) domain, which mediates protein interaction. Baculovirus IAPs also contain a copy of a really interesting new gene (RING) domain in the C-terminus [54]. Alphabaculoviruses carry various combinations of *iap-1*, *iap-2*, *iap-3*, and *iap-4*, whereas most, if not all, betabaculoviruses sequenced

to date contain *iap-3* and *iap-5*. Three members of *Gammabaculovirus* carry single *iap* homologs each that are most similar to *iap-3*, but encode only a single BIR and RING, or have two BIRs and lack a RING. Recently, the *iap-6* gene was identified in CpGV, CrleGV, PrGV, and PhopGV, all of which belong to *Betabaculovirus* [55]. In this study, we found not only *iap-3* (*orf11*) and *iap-5* (*orf105*) in the MaphGV genome, but also one *iap-6* (*orf85*). According to phylogenetic analysis, *iap* homologs in MaphGV are well separated, while *iap-6* from MaphGV is clustered with four other *iap-6* from CpGV, CrleGV, PrGV, and PhopGV (Figure 6). These GV's contain *iap-6* grouped in clade b of *Betabaculovirus* (Figure 3). To some extent, the phylogenies of *iap-6* and baculovirus genomes are coherent. Although the phylogeny of the five GV's containing *iap-6* is not strictly monophyletic, it is likely that the *iap-6* homolog descended from a common viral ancestor.

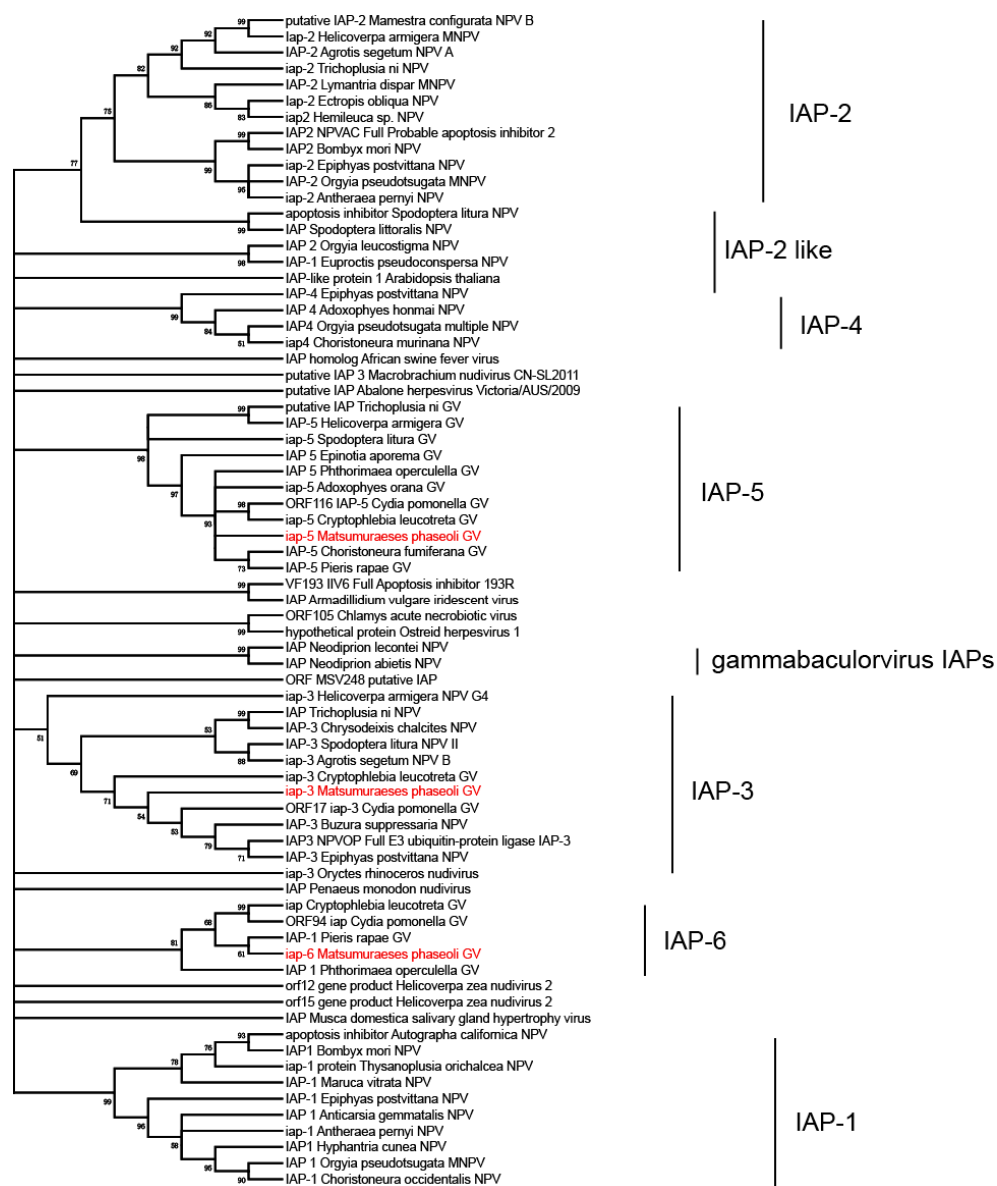


Figure 6. Phylogenetic analysis of the selected viral IAPs. The phylogenetic tree was constructed using the Maximum Likelihood method, using selected baculovirus IAP amino acid sequences with 500 bootstrap values. The initial tree for the heuristic search was obtained by applying the neighbor-joining method to a matrix of pairwise distances estimated using a JTT model. The sequences of IAPs are provided in Table S5. There were 1263 positions in the final dataset. The branches representing MaphGV IAPs are marked in red.

MaphGV *orf85* harbors one predicted BIR motif and a C-terminal RING domain, representing highly amino acid conservative property to other *iap-6* (Figure 7) [54]. Considering the phylogenetic relation and domain architecture, we inferred *orf85* of the MaphGV genome as the fifth *iap-6* in *Baculoviridae*. *iap-6* presumably lost a BIR domain during its evolutionary history [54]. Based on the above species phylogenetic analysis, viruses harboring *iap-6* were once closely related. The loss of the BIR domain may have occurred in their common ancestor.

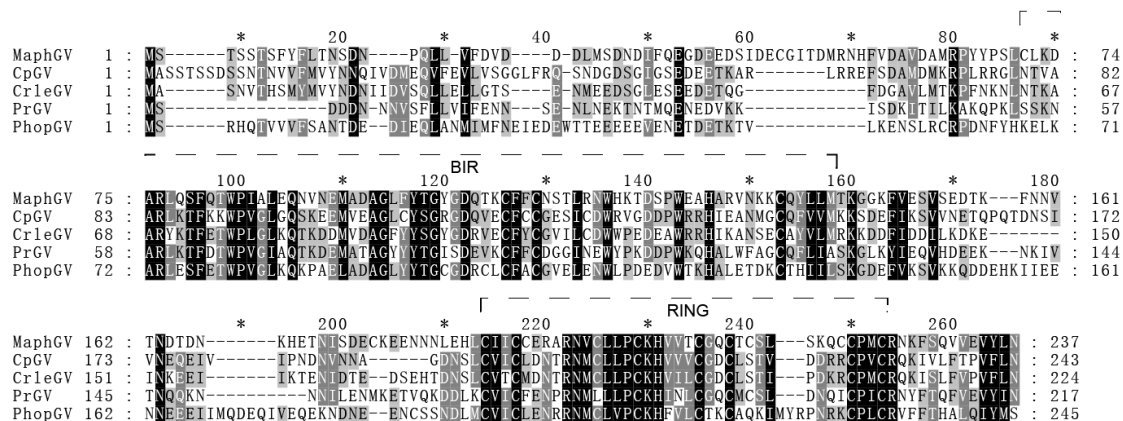


Figure 7. Predicted amino acid sequence alignment of IAP-6. The IAP-6 sequences of *Phthorimaea operculella granulovirus* (NP663251.1), CrleGV (NP891935.1), PrGV (YP003429403.1), CpGV (AIU37301.1), and MaphGV (this study) were aligned using MAFFT with default parameters. The identical amino acids are presented in white with a dark background, and the conserved ones are presented with a grey background. The BIR (71–142 aa) and RING (188–224 aa) domains of MaphGV *iap-6* are indicated by the dashed lines.

4. Conclusions

In this work, we described the genome of a baculovirus isolated from *M. phaseoli*. The virus, named MaphGV, is a novel species, phylogenetically clustered into clade b of the genus *Betabaculovirus*. The virus harbors 128 ORFs, of which 38 are baculovirus core genes, 18 are lepidopteran baculovirus conserved genes, and only 5 are unique to the family *Baculoviridae*. MaphGV is closely related to CpGV and CrleGV, displaying high collinearity with these related species. Two atypical *hrs* were found to be likely form long hairpin-loop structures. Sequencing and profiling of the MaphGV genome may be fundamental in bioinsecticides development. Thus, the discovery and genomic description of the novel baculovirus may guide the development of greener and safer pesticides to counteract and effectively control crop pest populations. These findings have enhanced our understanding on the evolution of baculovirus from a broader perspective.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1999-4915/12/10/1068/s1>, Figure S1: SNP of the sequenced MaphGV isolate, Table S1: Primers used in this study, Table S2: Core genes of 108 sequenced baculovirus genomes, Table S3: K2P pairwise distances of selected GVs, Table S4: *drs* in MaphGV, Table S5: *iap* genes of selected species.

Author Contributions: Conceptualization, Q.Q. and H.Z.; Data curation, R.S. and H.Z.; Formal analysis, R.S., Q.M., L.C. and W.J.; Funding acquisition, H.L., J.C. and Q.Q.; Investigation, L.M., H.L., J.C., Y.X. and Q.Q.; Methodology, R.S., Q.M. and H.L.; Project administration, Q.Q. and H.Z.; Resources, L.M., J.C. and Q.Q.; Software, R.S. and W.J.; Supervision, Q.Q. and H.Z.; Validation, L.C.; Visualization, R.S.; Writing—original draft, R.S.; Writing—review & editing, Q.M., Q.Q. and H.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Project of Science-Technology Basic Condition Platform from the Ministry of Science and Technology of the People’s Republic of China (Grant No. 2005DKA21402).

Acknowledgments: We thank Chunli Li from the Public Technology Service Center, Institute of Microbiology, Chinese Academy of Sciences for SEM assistance and preparation of Figure 1B.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Harrison, R.L.; Herniou, E.A.; Jehle, J.A.; Theilmann, D.A.; Burand, J.P.; Becnel, J.J.; Krell, P.J.; van Oers, M.M.; Mowery, J.D.; Bauchan, G.R.; et al. ICTV Virus Taxonomy Profile: *Baculoviridae*. *J. Gen. Virol.* **2018**, *99*, 1185–1186. [[CrossRef](#)]
2. Kelly, B.J.; King, L.A.; Possee, R.D. Introduction to Baculovirus Molecular Biology. In *Baculovirus and Insect Cell Expression Protocols*; Murhammer, D.W., Ed.; Methods in Molecular Biology; Springer: New York, NY, USA, 2016; pp. 25–50. ISBN 978-1-4939-3043-2.
3. Jehle, J.A.; Blissard, G.W.; Bonning, B.C.; Cory, J.S.; Herniou, E.A.; Rohrmann, G.F.; Theilmann, D.A.; Thiem, S.M.; Vlak, J.M. On the classification and nomenclature of baculoviruses: A proposal for revision. *Arch. Virol.* **2006**, *151*, 1257–1266. [[CrossRef](#)] [[PubMed](#)]
4. Clem, R.J.; Passarelli, A.L. Baculoviruses: Sophisticated Pathogens of Insects. *PLoS Pathog.* **2013**, *9*, e1003729. [[CrossRef](#)]
5. Herniou, E.A.; Olszewski, J.A.; O'Reilly, D.R.; Cory, J.S. Ancient Coevolution of Baculoviruses and Their Insect Hosts. *J. Virol.* **2004**, *78*, 3244–3251. [[CrossRef](#)] [[PubMed](#)]
6. Rohrmann, G.F. *Baculovirus Molecular Biology*, 4th ed.; National Center for Biotechnology Information (US): Bethesda, MD, USA, 2019.
7. Capinera, J.L. (Ed.) *Encyclopedia of Entomology*, 2nd ed.; Springer reference; Springer Science+Business Media B.V: Dordrecht, The Netherlands, 2008; ISBN 978-1-4020-6242-1.
8. Liu, P.; Zhao, H.; Luo, Y. Anti-Aging Implications of *Astragalus Membranaceus* (Huangqi): A Well-Known Chinese Tonic. *Aging Dis.* **2017**, *8*, 868–886. [[CrossRef](#)] [[PubMed](#)]
9. Byun, B.-K.; Park, K.-T.; Park, Y.-M. Review of the Genus *Matsumuraeses* Issiki (Lepidoptera, Tortricidae) with Discovery of *M. falcana* (Walsingham) in Korea. *J. Asia-Pac. Entomol.* **2005**, *8*, 117–122. [[CrossRef](#)]
10. Lü, H. Preliminary studies on the soy-bean leaf-roller *Matsumuraeses phaseoli* (Mats.), (Lepidoptera: Olethreutidae) I. morphology and biological characteristics. *Acta Phytophylacica Sin.* **1965**, *4*, 255–269. (In Chinese) [[CrossRef](#)]
11. Martignoni, M.E. *A Catalog of Viral Diseases of Insects, Mites, and Ticks*; U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station: Corvallis, OR, USA, 1986.
12. Hull, R.; Brown, F.; Payne, C. *Virology: A Directory and Dictionary of Animal, Bacterial and Plant Viruses*; Springer: Berlin, Germany, 1989; ISBN 978-1-349-07945-2.
13. Hajibabaei, M.; Janzen, D.H.; Burns, J.M.; Hallwachs, W.; Hebert, P.D.N. DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 968–971. [[CrossRef](#)]
14. Pinto, F.A.; Mattos, M.V.V.; Silva, F.W.S.; Rocha, S.L.; Elliot, S.L. The Spread of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and Coexistence with *Helicoverpa zea* in Southeastern Brazil. *Insects* **2017**, *8*, 87. [[CrossRef](#)]
15. Ratnasingham, S.; Hebert, P.D.N. BLOD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol. Ecol. Notes* **2007**, *7*, 355–364. [[CrossRef](#)]
16. Seo, B.Y.; Jung, J.K.; Cho, J.R.; Kim, Y.; Park, C.G. A PCR method to distinguish *Matsumuraeses phaseoli* from *M. falcana* based on the difference of nucleotide sequence in the mitochondrial cytochrome c oxidase subunit I. *Korean J. Appl. Entomol.* **2012**, *51*, 365–370. [[CrossRef](#)]
17. Tang, P.; Zhang, H.; Li, Y.; Han, B.; Wang, G.; Qin, Q.; Zhang, Z. Genomic sequencing and analyses of HearMNPV—a new Multinucleocapsid nucleopolyhedrovirus isolated from *Helicoverpa armigera*. *Virol. J.* **2012**, *9*, 168. [[CrossRef](#)] [[PubMed](#)]
18. Ardisson-Araújo, D.M.P.; de Melo, F.L.; de Souza Andrade, M.; Sihler, W.; Bão, S.N.; Ribeiro, B.M.; de Souza, M.L. Genome sequence of *Erinnyis ello granulovirus* (ErelGV), a natural cassava hornworm pesticide and the first sequenced sphingid-infecting betabaculovirus. *BMC Genomics* **2014**, *15*, 856. [[CrossRef](#)]
19. Sambrook, J.; Russell, D.W. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2001; ISBN 978-0-87969-577-4.
20. Hunt, M.; Gall, A.; Ong, S.H.; Brener, J.; Ferns, B.; Goulder, P.; Nastouli, E.; Keane, J.A.; Kellam, P.; Otto, T.D. IVA: Accurate de novo assembly of RNA virus genomes. *Bioinformatics* **2015**, *31*, 2374–2376. [[CrossRef](#)] [[PubMed](#)]
21. Wennmann, J.T.; Fan, J.; Jehle, J.A. Bacsnp: Using Single Nucleotide Polymorphism (SNP) Specificities and Frequencies to Identify Genotype Composition in Baculoviruses. *Viruses* **2020**, *12*, 625. [[CrossRef](#)]

22. Solovyev, V.; Kosarev, P.; Seledsov, I.; Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **2006**, *7*, S10. [CrossRef]
23. Wang, S.; Sundaram, J.P.; Spiro, D. VIGOR, an annotation program for small viral genomes. *BMC Bioinformatics* **2010**, *11*, 451. [CrossRef]
24. Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636–W641. [CrossRef]
25. Hu, Z.H.; Arif, B.M.; Jin, F.; Martens, J.W.M.; Chen, X.W.; Sun, J.S.; Zuidema, D.; Goldbach, R.W.; Vlak, J.M. Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **1998**, *79*, 2841–2851. [CrossRef]
26. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [CrossRef]
27. Kurtz, S.; Choudhuri, J.V.; Ohlebusch, E.; Schleiermacher, C.; Stoye, J.; Giegerich, R. REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* **2001**, *29*, 4633–4642. [CrossRef] [PubMed]
28. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef] [PubMed]
29. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276–277. [CrossRef]
30. Lorenz, R.; Bernhart, S.H.; Höner zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26. [CrossRef] [PubMed]
31. Liu, W.; Xie, Y.; Ma, J.; Luo, X.; Nie, P.; Zuo, Z.; Lahrmann, U.; Zhao, Q.; Zheng, Y.; Zhao, Y.; et al. IBS: An illustrator for the presentation and visualization of biological sequences. *Bioinformatics* **2015**, *31*, 3359–3361. [CrossRef] [PubMed]
32. BoxShade. Available online: https://embnet.vital-it.ch/software/BOX_form.html (accessed on 15 July 2020).
33. Shen, W.; Le, S.; Li, Y.; Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS ONE* **2016**, *11*, e0163962. [CrossRef]
34. Guarino, L.A.; Smith, M. Regulation of delayed-early gene transcription by dual TATA boxes. *J. Virol.* **1992**, *66*, 3733–3739. [CrossRef]
35. Blissard, G.W.; Kogan, P.H.; Wei, R.; Rohrmann, G.F. A synthetic early promoter from a baculovirus: Roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. *Virology* **1992**, *190*, 783–793. [CrossRef]
36. Wormleaton, S.; Kuzio, J.; Winstanley, D. The complete sequence of the *Adoxophyes orana* granulovirus genome. *Virology* **2003**, *311*, 350–365. [CrossRef]
37. Morris, T.D.; Miller, L.K. Mutational analysis of a baculovirus major late promoter. *Gene* **1994**, *140*, 147–153. [CrossRef]
38. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
39. Criscuolo, A.; Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **2010**, *10*, 210. [CrossRef] [PubMed]
40. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [CrossRef] [PubMed]
41. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **1980**, *16*, 111–120. [CrossRef] [PubMed]
42. Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinforma. Oxf. Engl.* **2004**, *20*, 289–290. [CrossRef] [PubMed]
43. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef] [PubMed]
44. King, A.M.Q.; Adams, M.J. *Virus Taxonomy Ninth Report of the International Committee on Taxonomy of Viruses*; Elsevier Science & Technology Books: San Diego, CA, USA, 2011; ISBN 978-0-12-384685-3.

45. Garavaglia, M.J.; Miele, S.A.B.; Iserte, J.A.; Belaich, M.N.; Ghiringhelli, P.D. The ac53, ac78, ac101, and ac103 Genes Are Newly Discovered Core Genes in the Family Baculoviridae. *J. Virol.* **2012**, *86*, 12069–12079. [[CrossRef](#)] [[PubMed](#)]
46. Jehle, J.A.; Lange, M.; Wang, H.; Hu, Z.; Wang, Y.; Hauschild, R. Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. *Virology* **2006**, *346*, 180–193. [[CrossRef](#)] [[PubMed](#)]
47. Wennmann, J.T.; Keilwagen, J.; Jehle, J.A. Baculovirus Kimura two-parameter species demarcation criterion is confirmed by the distances of 38 core gene nucleotide sequences. *J. Gen. Virol.* **2018**, *99*, 1307–1320. [[CrossRef](#)] [[PubMed](#)]
48. Herniou, E.A.; Olszewski, J.A.; Cory, J.S.; O'Reilly, D.R. The genome sequence and evolution of baculoviruses. *Annu. Rev. Entomol.* **2003**, *48*, 211–234. [[CrossRef](#)] [[PubMed](#)]
49. Hilton, S.; Winstanley, D. The origins of replication of granuloviruses. *Arch. Virol.* **2008**, *153*, 1527–1535. [[CrossRef](#)] [[PubMed](#)]
50. Harrison, R.L.; Rowley, D.L.; Mowery, J.; Bauchan, G.R.; Theilmann, D.A.; Rohrmann, G.F.; Erlandson, M.A. The Complete Genome Sequence of a Second Distinct Betabaculovirus from the True Armyworm, *Mythimna unipuncta*. *PLoS ONE* **2017**, *12*, e0170510. [[CrossRef](#)] [[PubMed](#)]
51. Hayakawa, T.; Ko, R.; Okano, K.; Seong, S.I.; Goto, C.; Maeda, S. Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* **1999**, *262*, 277–297. [[CrossRef](#)] [[PubMed](#)]
52. Cuartas, P.E.; Barrera, G.P.; Belaich, M.N.; Barreto, E.; Ghiringhelli, P.D.; Villamizar, L.F. The Complete Sequence of the First *Spodoptera frugiperda* Betabaculovirus Genome: A Natural Multiple Recombinant Virus. *Viruses* **2015**, *7*, 394–421. [[CrossRef](#)]
53. Kang, W.; Suzuki, M.; Zemskov, E.; Okano, K.; Maeda, S. Characterization of Baculovirus Repeated Open Reading Frames (*bro*) in *Bombyx mori* Nucleopolyhedrovirus. *J. Virol.* **1999**, *73*, 10339–10345. [[CrossRef](#)]
54. Clem, R.J. Viral IAPs, then and now. *Semin. Cell Dev. Biol.* **2015**, *39*, 72–79. [[CrossRef](#)]
55. Harrison, R.L.; Rowley, D.L.; Funk, C.J. The Complete Genome Sequence of Plodia Interpunctella Granulovirus: Evidence for Horizontal Gene Transfer and Discovery of an Unusual Inhibitor-of-Apoptosis Gene. *PLoS ONE* **2016**, *11*, e0160389. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).