

Article

Sequencing, Analysis and Organization of the Complete Genome of a Novel Baculovirus *Calliteara abietis* Nucleopolyhedrovirus (CaabNPV)

Wenyi Jin ^{1,2}, Mijidsuren Byambasuren ³, Uranbileg Ganbold ^{1,2,3}, Huixian Shi ¹, Hongbin Liang ⁴, Miaomiao Li ⁵, Hongtuo Wang ¹, Qilian Qin ¹ and Huan Zhang ^{1,*} 

¹ State Key Laboratory of Integrated Management of Pests and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; jinwenyi23@ioz.ac.cn (W.J.); uranbileg@ioz.ac.cn (U.G.); s18348965602@163.com (H.S.); wanght@ioz.ac.cn (H.W.); qinql@ioz.ac.cn (Q.Q.)

² University of Chinese Academy of Sciences, Beijing 101408, China

³ Institute of Plant Protection, Mongolian University of Life Science, Ulaanbaatar 627153, Mongolia; bybyambasuren4747@gmail.com

⁴ Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; lianghb@ioz.ac.cn

⁵ Institute of College of Basic Medicine, Shaanxi University of Chinese Medicine, Xianyang 712046, China; 1351029@sntcm.edu.cn

* Correspondence: zhanghuan@ioz.ac.cn; Tel.: +86-10-6480-7262

Abstract: *Baculoviridae*, a virus family characterized by a single large double stranded DNA, encompasses the majority of viral bioinsecticides, representing a highly promising and environmentally friendly pesticide approach to insect control. This study focuses on the characterization of a baculovirus isolated from larvae of *Calliteara abietis* (Erebidae, Lymantriidae) collected in Mongolian pinaceae forests. This new isolate was called *Calliteara abietis* nucleopolyhedrovirus (CaabNPV). CaabNPV exhibits an irregular polyhedron shape, and significant variation in the diameter of its occlusion bodies (OBs) was observed. Nucleotide distance calculations confirmed CaabNPV as a novel baculovirus. The CaabNPV genome spans 177,161 bp with a G+C content of 45.12% and harbors 150 potential open reading frames (ORFs), including 38 core genes. A comprehensive genomic analysis categorizes CaabNPV within Group II alphabaculovirus, revealing a close phylogenetic relationship with *Alphabaculovirus orleucostigmae* (OrleNPV). Additionally, repeat sequence analysis identified three highly repetitive sequences consisting of 112 bp repeat units, known as homologous regions (*hrs*). This research contributes valuable insights into CaabNPV's phylogenetic placement, genomic structure, and its potential applications in insect biocontrol.

Keywords: *Alphabaculovirus*; baculovirus genome; *Calliteara abietis*



Citation: Jin, W.; Byambasuren, M.; Ganbold, U.; Shi, H.; Liang, H.; Li, M.; Wang, H.; Qin, Q.; Zhang, H. Sequencing, Analysis and Organization of the Complete Genome of a Novel Baculovirus *Calliteara abietis* Nucleopolyhedrovirus (CaabNPV). *Viruses* **2024**, *16*, 252. <https://doi.org/10.3390/v16020252>

Academic Editor: Miguel López-Ferber

Received: 25 December 2023

Revised: 31 January 2024

Accepted: 3 February 2024

Published: 4 February 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the current context, approximately one hundred species of baculoviruses have been identified, with a focus on harnessing their insecticidal potential for protecting field crops [1]. However, the prevailing challenge lies in the host specificity exhibited by the majority of these baculoviruses, underscoring the urgent need for ongoing exploration and the development of novel viral strains. Simultaneously, the discovery of new baculovirus holds paramount significance, not only for advancing crop protection strategies but also for garnering essential insights into their evolution and intricate interactions with host organisms [2]. This exploration not only enhances the arsenal of crop protection methodologies but also contributes significantly to our understanding of the dynamic interplay between viruses and their hosts.

Baculoviridae, a family of large double-stranded DNA viruses, emerges as a promising and environmentally friendly pesticide with the capability to address the threats posed by

insects across various orders, including Lepidoptera, Hymenoptera, Diptera, and others [3]. Characterized by a single circular double-stranded DNA ranging from 80 to 180 kb [4] and encoding 80 to 190 genes [4], the genome is enclosed within rod-shaped nucleocapsids measuring 230 to 385 nm in length and 40 to 60 nm in diameter [5]. The distinctive life cycle of *Baculoviridae* involves the production of two progeny viruses with divergent structures and functions. One viral phenotype, referred to as budded virus (BV), facilitates systematic infection spread between cells within a host, while the other, known as occlusion-derived virus (ODV), transmits infection from insect to insect via the oral route [6]. This ancient virus demonstrates the capacity for co-evolution with its original host [7,8]. One defining characteristic of baculoviruses is the presence of a conserved set of 38 genes which form a shared core that spans their entire genomic structure. The classification of baculoviruses into four distinct genera: *Alphabaculovirus* (lepidopteran-specific nucleopolyhedroviruses [NPVs]), *Betabaculovirus* (lepidopteran-specific granuloviruses [GVs]), *Gammabaculovirus* (hymenopteran-specific NPVs), and *Deltabaculovirus* (dipteran-specific NPVs), respectively. Additionally, *Alphabaculovirus* further divides into Group I and Group II [4]. Group I alphabaculoviruses utilize GP64 as their envelope fusion protein (EFP) in the BV, while Group II alphabaculoviruses lack GP64 and instead employ F protein as their EFP [9].

Calliteara abietis (Erebidae: Lymantriidae), a significant pest of Pinaceae, is prevalent in northern and eastern Europe, but has limited presence in northern Asia. Targeting primarily *Pinus sylvestris*, larch trees, and fir trees, this pest causes extensive damage to pine forests across vast territories [10]. Its wide distribution and destructive feeding habits underscore its impact on forestry ecosystems in these regions.

Genome research serves as a crucial source of illumination, unveiling the enigmatic facets of baculoviruses. Through meticulous analysis, baculoviruses can be precisely identified and classified, revealing the intricacies of their functional genes and encoded proteins. Beyond providing new theoretical foundations for virus control, this study initiates a comprehensive exploration into the biology of the virus and its intricate interactions with host organisms. In doing so, it establishes a robust foundation that not only deepens our current understanding but also serves as a springboard for future research endeavors. However, given the narrow host specificity of most baculoviruses, the discovery and sequence analysis of new baculoviruses remain crucial.

In this study, the Erebidae larvae exhibiting signs of baculoviruses infection were collected from a forest in Mongolia. Through a combination of morphological characteristics and mitochondrial cytochrome oxidase I (COI) sequence data, the host was identified as *C. abietis*. The baculovirus isolated from the larvae underwent a complete genome analysis, designated as *Calliteara abietis* nucleopolyhedrovirus (CaabNPV). Comparative analysis with other fully sequenced baculovirus genomes confirmed that CaabNPV represents a novel virus.

2. Materials and Methods

2.1. Baculovirus Isolation and Host Identification

CaabNPV and its host larva were collected from Bayan-Adraga, Mongolia (coordinates: 48°32'14.5" N, 111°04'44.5" E, altitude 1034 m) in 2023. *C. abietis* larvae were collected, including uninfected individuals and individuals showing signs of infection.

Insect DNA samples were extracted using the MightyPrep Reagent for DNA (TAKARA). The amplification of the COI barcode was conducted using primers specifically designed for the Erebidae [11]. The amplified fragments were subjected to Sanger sequencing, and the obtained sequences were compared and analyzed against the barcode sequences available in GenBank [12]. Primer sequence information can be found in Table S1.

The OBs were derived from the larvae cadavers that succumbed to the infection, and purification of occlusion bodies (OBs) was performed through a series of steps, including maceration, homogenization, filtration, centrifugation, and sucrose density-gradient ultracentrifugation [13].

2.2. Electron Microscopy

Baculovirus-specific characteristics were examined utilizing light microscopy (Leica DM2000, Leica, Wetzlar, Germany) to assess the purity of the baculovirus.

For a more in-depth exploration of structural nuances, purified OBs underwent detailed observation using a scanning electron microscope (SEM, Hitachi, Tokyo, Japan) with a 5 kV acceleration voltage. The dimensions of CaabNPV OBs were precisely ascertained, relying on measurements extracted from a comprehensive set of 69 complete SEM image samples.

2.3. Viral DNA Isolation, Sequencing, and Assembly

The purified OBs was resuspended in an alkaline solution (100 mM Na₂CO₃, pH = 10.8) and incubated at 37 °C for 30 min. The pH of the solution was adjusted to 8.0 using 1M hydrochloric acid. Ribonuclease A and Proteinase K was added to the solution and incubated, following the initial incubation at 37 °C for 2 hours, the solution was subjected to a subsequent incubation at 65 °C for an additional 2 hours. Viral DNA was extracted using a phenol/chloroform/isoamyl alcohol (25:24:1) mixture and ethanol-precipitated. The DNA concentration was determined using a spectrophotometer, and the purity was validated by the 260/280 ratio.

CaabNPV full genome sequencing was performed on DNB-seq. After data filtration, the average read length was determined to be 150 bp, and the achieved coverage reached 38,135×. The Q20 and Q30 contents were 96.14%, and 90.44%, respectively. Unicycler (version v0.5.0) was used to assemble high-quality pair-end reads into contigs by de novo assembly.

The A region with a low mapping rate was amplified through PCR and assembly errors were resolved through Sanger sequencing [14]. Additionally, CaabNPV exhibits an assembly ambiguity region. To identify this region, vIdentify-F/R primers (Table S1) were utilized for the amplification of both CaabNPV and *C. abietis*. To ensure that the experimental sample of *C. abietis* was not contaminated with baculovirus, specific primers lef8-F/R (Table S1) were employed to amplify the baculovirus for host larvae identification.

The amplification process followed the manufacturer's instructions using TransStart® FastPfu DNA Polymerase (Beijing, TransGen Biotech Co., Ltd., Beijing, China). The reaction conditions were as follows: initial denaturation at 95 °C for 2 minutes, followed by 35 cycles of denaturation at 95 °C for 20 seconds, annealing at 55 °C for 20 seconds, and extension at 72 °C for 30 seconds. A final extension was performed at 72 °C for 5 minutes.

2.4. Genome Sequence Analysis

Open reading frames (ORFs) encoding proteins with a minimum of 50 amino acids were predicted using both FGENESV [15] and the NCBI ORF Finder. The annotation convention for overlapping ORFs, limited to 75 base pairs, involves annotating only the larger ORF, unless the smaller one corresponds to a conserved baculovirus homolog [16]. This approach stems from the assumption that the expression of the smaller ORF is less probable. ORFs within homologous region (*hr*) sequences, due to their relative instability, were left unannotated.

The identified ORFs were annotated based on homology using the Protein-Protein BLAST algorithm. The entire genomic data were submitted to Genbank under accession number PP171514.

Gene parity plots were generated to compare ORF organization between CaabNPV and the selected baculoviruses [17].

EMBOSS stretcher [18] was employed to assess the genomic homology between CaabNPV and other baculoviruses. *hrs* were identified using the tandem repeats finder [19], and EMBOSS palindrome [20] was utilized to search for palindrome sequences. The ViennaRNA secondary structure server [21] was applied to predict the secondary structure of the conserved *hr* sequence.

To identify early and late promoter motifs, a 180-nucleotide upstream region of each initiation codon was screened. The criteria for identifying the early promoter component included the presence of a common TATA-box motif (TATAW) with a CAKT mRNA start site sequence located 25–35 nucleotides downstream [22,23]. Moreover, elements similar to TATA, such as (TAATWAA), were identified upstream of early expression factors [24]. The late promoter component was defined as DTAAG, typically located around -60 nucleotides upstream of the ORF initiation codon [25].

2.5. Phylogeny and Kimura 2-Parameter Analysis

Protein sequences encoded by the 38 core genes were extracted from 108 baculovirus genomes, with inclusion of CaabNPV (Table S2), and aligned using the MAFFT method with auto strategy and normal alignment mode [26]. The alignment was performed using OrthoFinder (v2.5.5) with default parameters [27]. Following sequence alignment, the genomes were concatenated while maintaining the order of each gene. The optimal partitioning scheme for maximum likelihood (ML) was determined using RAxML-NG (v1.2.0) [28]. The best partitioning scheme for ML was identified by PartitionFinder2 using the Bayesian information criterion (BIC) and a greedy search algorithm with branch lengths linked [29]. The ML tree was inferred with IQ-TREE under Ultrafast bootstrap with 5000 replicates.

Kimura 2-parameter (K2P) pairwise distances [30] from aligned nucleotide sequences of *polyhedrin/lef-8/lef-9* fragments were calculated separately using the pairwise distance calculation in PHYLIP (v 3.6). Substitution rates among sites were set to be uniform. Gaps within the alignment were treated as missing data, ensuring a comprehensive analysis of sequence variations.

3. Results and Discussion

3.1. Host Determination and Virus Characterization

In Mongolia, viruses and their host larva were collected in the field (Figure 1a,b). Based on larval morphology, they were classified as a larvae species in *Calliteara* genus (Erebidae: Lymantriidae). The deceased larvae samples collected on pine trees exhibited pathological features associated with baculovirus infection (Figure 1b). These characteristics included body darkening, liquefaction, and hanging upside down on branches. Sequencing of the mitochondrial COI barcode (GenBank accession number: KX436516.1) confirmed the host as *C. abietis*. The sequencing results can be found in Appendix A. Consequently, the virus was named *Calliteara abietis* nucleopolyhedrovirus (CaabNPV).

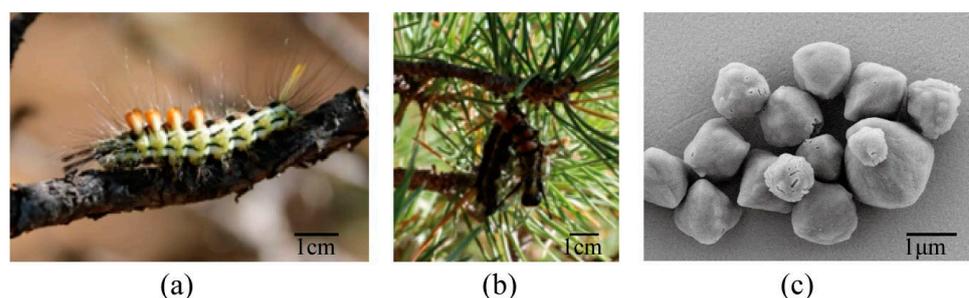


Figure 1. *Calliteara abietis* larva and *Calliteara abietis* nucleopolyhedrovirus (CaabNPV) scanning electron microscope (SEM). (a) A *C. abietis* larva feeding on pine tree. (b) Deceased *C. abietis* displaying pathological features indicative of baculovirus infection. (c) SEM picture depicting irregularly shaped CaabNPV occlusion bodies (OBs).

Under optical microscopy, polyhedral of OBs displaying Brownian motion were observed. SEM analysis of CaabNPV unveiled an ultrastructure characterized by irregularly shaped polyhedra (Figure 1c), indicative of a typical feature observed in OBs of NPV. Quantification of individual OBs demonstrated a mean diameter of approximately $2.19 \pm 0.56 \mu\text{m}$, falling within the size range previously reported for baculoviruses in the

literature [5]. The size variation observed in CaabNPV OBs may be attributed to its natural collection, indicating that the strain has not undergone purification to establish a uniform viral population. This observation aligns with established findings, further supporting the classification of CaabNPV within the *Alphabaculovirus* genus.

3.2. Nucleotide Sequence Analysis

Utilizing the DNB-seq high-throughput sequencing platform, a total of 4,023,799 high-quality pair-end reads were generated for the CaabNPV sample. The complete genome was assembled using Unicycler v0.5.0. Subsequently, validation of ambiguous regions was conducted through PCR and Sanger sequencing methods.

The analysis of the assembled results for the ambiguous region revealed the presence of two variant assemblies. As a result of the baculovirus being purified before sequencing, any potential host fragments were effectively removed, indicating that both identified segments originated from CaabNPV. The region is located downstream of the *polh* initiation site at 16,792 bp. One variant comprises a 189 bp repetitive sequence interspersed with a 793 bp contig, while the other variant consists of a single 189 bp contig (Figure 2a, with sequence details provided in Appendix B). PCR and Sanger sequencing verification using vIdentify-F/R primers further confirmed the simultaneous existence of these two variants within CaabNPV (Figure 2b). To eliminate the possibility of host contamination, samples were collected from wild host larvae without baculovirus infection. Verification using baculovirus gene primers *lef8*-F/R [31] confirmed the absence of CaabNPV infection in the host (Figure 2c lane 3, positive control in Figure 2c lane 4). Subsequent amplification of host DNA using vIdentify-F/R primers resulted in the generation of two expected bands, as depicted in Figure 2c lane 2. Further sequencing results indicated the presence of both the 189 bp and 1171 bp variants in the uninfected host genome, affirming the concurrent existence of these two variants within the host and CaabNPV genome.

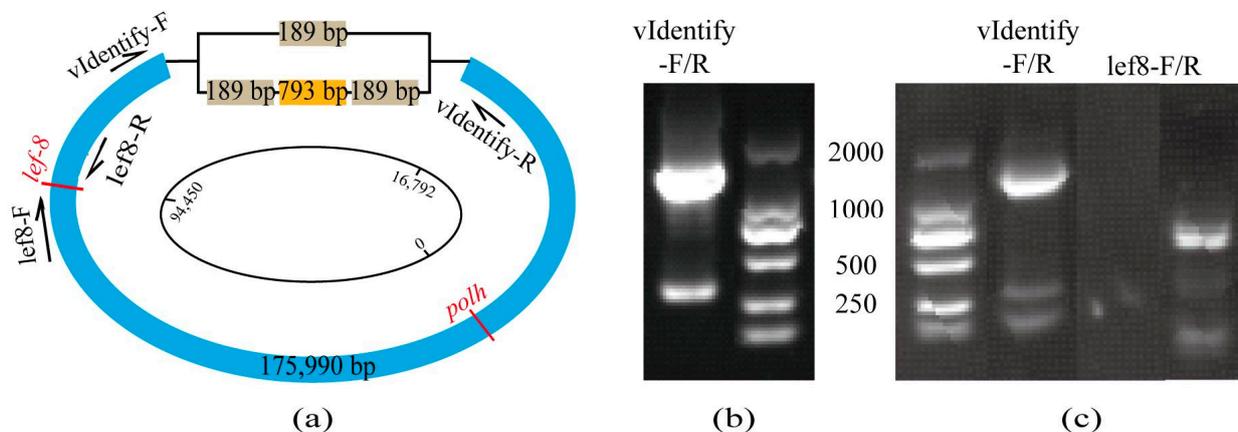


Figure 2. CaabNPV assembly results (a) The schematic diagram illustrates the assembly results of CaabNPV, emphasizing ambiguous regions. Arrows indicate primer sequences, and gene names are denoted in red, the light brown box represents a 189bp fragment, the orange box represents a 793bp fragment, and the remaining part of the CaabNPV genome is represented by the blue box. (b) PCR amplification of ambiguous regions in CaabNPV with vIdentify-F/R reveals large fragments consisting of two repeated 189 bp sequences interspersed with a 793 bp sequence, while small fragments consist of a singular 189 bp sequence. (c) Lane 2: Amplification of ambiguous regions in host larvae, depicting PCR amplification of ambiguous regions in the host larvae; Lane 3: Amplification of the viral *lef8* fragment in host larvae; Lane 4: *lef8* positive control with CaabNPV as the sample, resulting in an amplification size of 849 bp.

The presence of a transposon-like structure is likely to be attributed to horizontal gene transfer from the host genome during virus replication [32,33]. Additionally, the coexistence of two distinct genomes within a host organism was observed. This phenomenon may be attributed to the natural survival of a mixed population, comprising multiple genotypes, of an insect nucleopolyhedrovirus. Such diversity is likely to be maintained to facilitate the efficient transmission of the virus by insects. Infections involving multiple genotypes tend to generate OBs with higher pathogenicity compared to those produced by any single genotype alone [34].

However, due to the inability to artificially rear host larvae and obtain purified individual baculoviral strains, the determination of whether the ambiguous sequence results from horizontal gene transfer between the baculovirus and host or is due to the inherent presence of multiple genotypes within OBs is hindered. Despite inferring based on contig depth, with a depth of $2.01\times$ for the 189 bp contig and $0.98\times$ for the 793 bp contig, a comprehensive understanding of the virus structure and function remains constrained by the lack of detailed explanations for such transposon-like structures.

Furthermore, contig depth-based estimation revealed that the chosen assembly result for further analysis of the CaabNPV genome includes two repetitive 189 bp sequences interspersed with a 793 bp sequence (Figure 2a). Due to the absence of homologous gene matches and unreported host genome sequences, the current elucidation of this transposon-like structure remains inadequate. This limitation will impede a comprehensive understanding of virus structure and function.

The circular genome of CaabNPV measured 177,161 bp in size. The CaabNPV genome exhibited a G+C content of 45.12%, similar to the average GC content observed in group I (44.9%) and group II (41.6%) alphabaculoviruses [35]. Following the original definition, the polyhedrin gene was designated as ORF1, and its methionine start codon was defined as the zero point on the CaabNPV physical map (Figure 3).

In total, 150 putative ORFs were identified (Table S3), each with a minimum length of 50 amino acids. Remarkably, CaabNPV encoded a higher number of genes (150 ORFs) compared to other baculoviruses, where the average number of ORFs was 140.5 [36]. These ORFs collectively accounted for 83% of the entire genome. Among them, 83 (55.3%) ORFs were oriented in the forward direction, while 67 (44.6%) were in the reverse orientation (Figure 3). CaabNPV encompassed 38 core genes shared across all baculoviruses (Table S2). It also contains 23 genes that were conserved among baculoviruses [37] and had 72 common genes found in the *Baculoviridae* family.

In the upstream region preceding the start codon of the 17 unique genes in CaabNPV, spanning 180 base pairs, a comprehensive analysis was conducted. Within the 3 ORFs (Caab_082, Caab_138, Caab_139), early promoter sequences were identified, comprising TATA box motifs (TATAW and TAATWAA) and CAKT. The average distance between these motifs was determined to be 29.6 bp. Furthermore, eight ORFs (Caab_020, Caab_028, Caab_036, Caab_080, Caab_081, Caab_099, Caab_100, Caab_136) contained the late promoter element (DTAAG), which was positioned at an average distance of -68 bp upstream of the start codon. Notably, two ORFs (Caab_087, Caab_137) exhibited both early and late promoter motifs. Interestingly, a total of four ORFs (Caab_027, Caab_101, Caab_116, Caab_117) lacked recognizable typical promoter motifs. Further investigations are essential to determine the functional significance of these findings in CaabNPV.

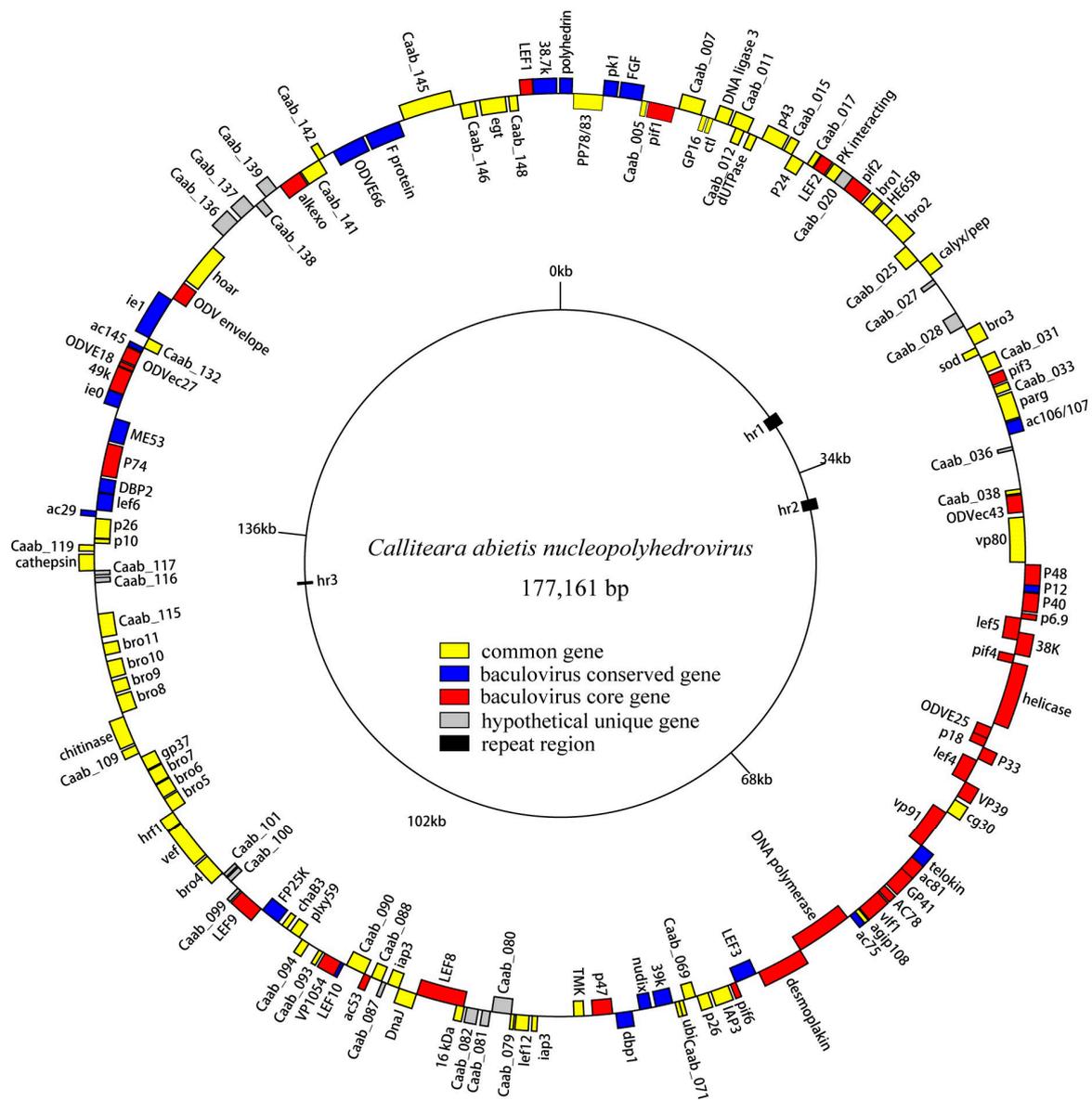


Figure 3. Circular map of CaabNPV genome. The colors represent gene types: red represents baculovirus core genes, blue represent baculovirus conserved genes, yellow represents other baculovirus common genes, gray represents hypothetical genes unique to CaabNPV, black represents repeat regions.

3.3. Phylogenetic Analysis

To investigate the evolutionary relationship between CaabNPV and other baculoviruses, a phylogenetic tree was constructed based on the amino acid sequences of 38 highly conserved core genes from 108 baculovirus genomes (Table S2). CaabNPV encodes the F protein and lacks the *gp64* gene (Figure 3), placing it within the Group II of genus *Alphabaculovirus*. In the ML phylogram, alphabaculoviruses clustered into a monophyletic group, and Group I alphabaculoviruses and Group II alphabaculoviruses formed a monophyletic clade within the alphabaculovirus group. This topological structure is in accordance with previous studies [9]. Based on the phylogenetic tree analysis, CaabNPV is classified within the Group II alphabaculovirus (Figure 4). CaabNPV is a sister species to *Alphabaculovirus orleucostigmae* (OrleNPV) and has a nucleotide identity of 47.2%.

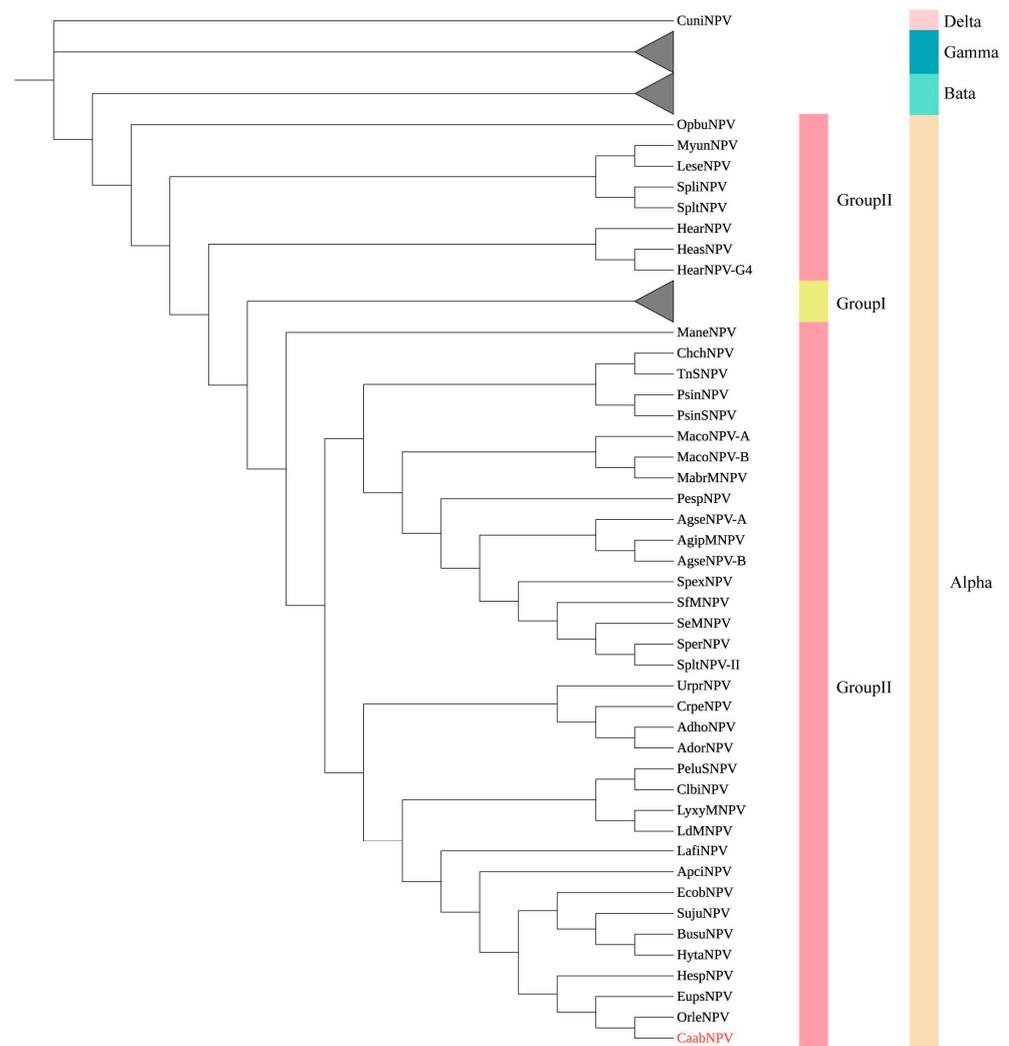


Figure 4. Unrooted phylogenetic tree created from the concatenated 38 core genes found in 108 sequenced baculovirus genomes. CaabNPV is marked in red. The tree was constructed using the minimum evolution method with 1000 bootstrap. Clusters representing the group I alphabaculovirus genera, *Beta-*, *Gammabaculovirus* have been simplified for clarity.

Based on the pairwise nucleotide distances estimated using the K2P model with *polyhedrin/granulin*, *lef-8*, and *lef-9* genes, whether CaabNPV was a novel baculovirus was identified [31]. The K2P pairwise for CaabNPV and other baculovirus above-mentioned genes were greater than 0.05 substitutions/site (Table S4), indicated that CaabNPV was a new baculovirus.

The gene parity plot was used to compare the gene order among different baculovirus genomes. Its purpose was to illustrate the collinearity among baculovirus genomes and the patterns of baculovirus evolution [17]. *Culex nigripalpus* NPV (CuniNPV, *Deltabaculovirus*, NC_003084), *Neodiprion sertifer* NPV (NeseNPV, *Gammabaculovirus*, NC_005905), *Adoxyphes orana* granulovirus (AdorGV, *Betabaculovirus*, NC_005038), *Autographa californica* nucleopolyhedrovirus (AcMNPV, group I alphabaculovirus, NC_001623), *Lymantria dispar* MNPV (LdMNPV, group II alphabaculovirus, NC_001973), were used as representative baculoviruses, and *OrleNPV* (NC_010276), *Hemileuca* sp. nucleopolyhedrovirus (HespNPV, NC_021923) and *Euproctis pseudoconsersa* nucleopolyhedrovirus (EupsNPV, NC_012639) were used as related baculoviruses. The gene order of CaabNPV was compared with the genomes mentioned above. CaabNPV shared 36 homologous ORFs with CuniNPV, 49 with NeseNPV, 77 with AdorGV, 105 with AcMNPV, 120 with LdMNPV, 112 with OrleNPV, 112 with HespNPV and 112 with EupsNPV.

The whole-genome nucleotide identity between the CaabNPV genome and those of OrleNPV, HespNPV, and EupsNPV was 47.2%, 44.1% and 44.4%, respectively. CaabNPV exhibited a highly collinear gene order with OrleNPV and EupsNPV, indicating a similarity in gene structure. The gene parity plots reveal that the majority of ORFs in OrleNPV and EupsNPV are arranged in an inverse order compared to the CaabNPV ORFs. This is attributed to the artificial definition of *polh* as the first ORF, leading to the occurrence of an inversion in the region encompassing the *polh*. In contrast, its gene arrangement significantly differed from that of NeseNPV (*Gammabaculovirus*), CuniNPV (*Deltabaculovirus*), and AdorGV (*Betabaculovirus*) (Figure 5).

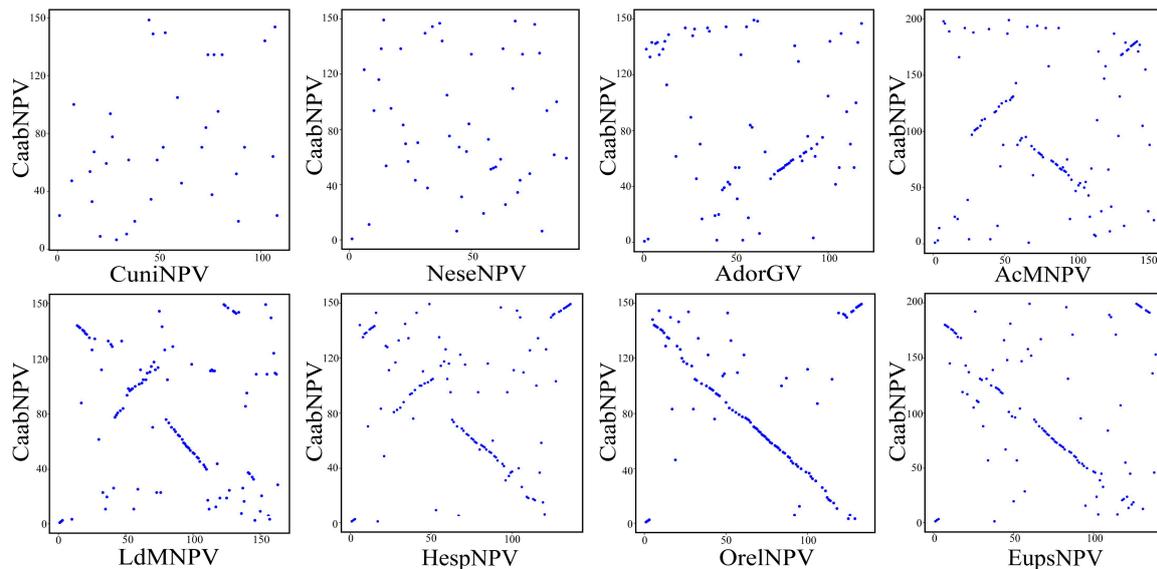


Figure 5. Gene parity plots comparing CaabNPV with other baculoviruses. The ORF content and order of the CaabNPV genome (y-axis) was compared with that of *Culex nigripalpus* NPV (CuniNPV, *Deltabaculovirus*), *Neodiprion sertifer* NPV (NeseNPV, *Gammabaculovirus*), *Adoxophyes orana* granulovirus (AdorGV, *Betabaculovirus*), *Autographa californica* nucleopolyhedrovirus (AcMNPV, group I alphabaculovirus), *Lymantria dispar* MNPV (LdMNPV, group II alphabaculovirus) and *Orygia leucostigma* NPV (OrleNPV), *Hemileuca* sp. nucleopolyhedrovirus (HespNPV) and *Euproctis pseudoconspersa* nucleopolyhedrovirus (EupsNPV) (x-axes). Each point in the plot represented the relative position of open reading frames in the CaabNPV genome with those from other baculoviruses.

3.4. Repeated Sequences

Most of baculovirus *hrs* contain repeated AT-rich sequences [38]. The *hrs* are characterized by comprising a series of tandem repeats, which include an imperfect palindromic core, interspersed at different locations in the genome [16]. The *hrs* play putative or demonstrated roles as origins of DNA replication and enhancers of baculovirus early promoters [39].

The CaabNPV genome contained three *hrs*, and each consisted of between two and four repeat units of approximately 112 bps. Near the repeat unit center, there was an imperfect palindrome spanning 26 bps (Figure 6a,b). As depicted in Figure 3, *hr1* is located between Caab_27 and Caab_28, and *hr2* between Caab_37 and Caab_38; they contain four and three repeat units, respectively. *hr3* was positioned between Caab_118 and Caab_119, featuring a tandem repeat of two copies. Secondary structure predictions were illustrated in Figure 6c, and the *hr1* contained four repeating units, showing a more complex secondary structure.

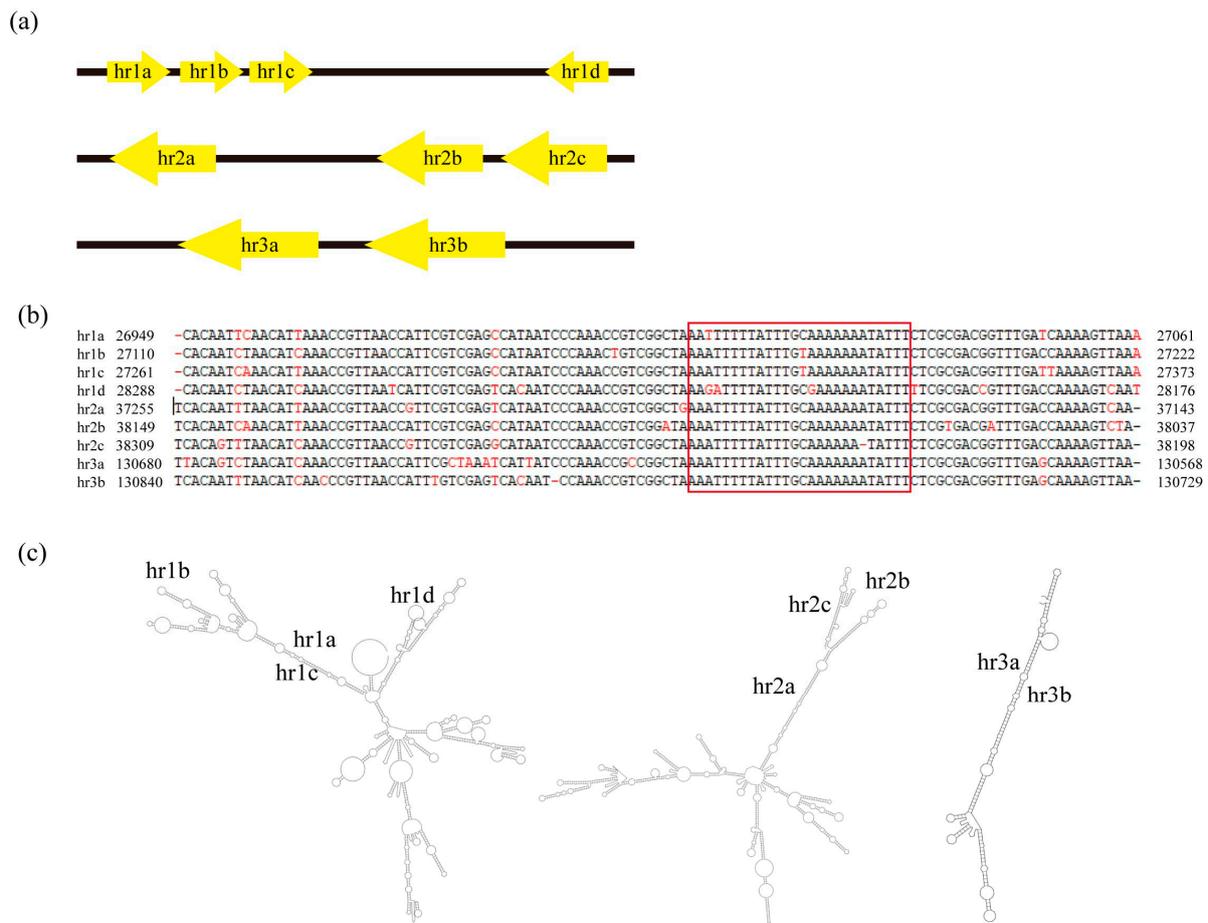


Figure 6. Analysis of CaabNPV homologous regions (*hrs*). (a) The schematic diagram for location and structure of the five *hrs* in the CaabNPV genome. (b) Sequence alignment of CaabNPV *hrs* repeat units. The imperfect palindromic sequence was represented with Group II alphabaculoviruses in the red box. (c) Secondary structure prediction of the conserved *hrs* sequence.

3.5. Analysis of CaabNPV Multi-Copy Genes

The majority of sequenced baculovirus genomes exhibit a range of 1–16 copies of the *bro*, and its function is not clear yet [40]. The Bro protein possesses DNA-binding capabilities and is associated with the modulation of host DNA replication in *Alphabaculovirus bomori* (BmNPV) [41]. In CaabNPV, there were a total of 11 Bro proteins, among which multiplied Bros were adjacent (Caab_105- Caab_107 and Caab_111- Caab_114).

Except for CuniNPV, homologs of the DNA-binding protein (DBP) are evident in all cataloged baculovirus genomes. Multiple copies of DBP genes were observed in Group II alphabaculoviruses [42]. The DBP gene plays a role in the unwinding and annealing of DNA during the replication process [43]. Notably, the CaabNPV genome contained two copies of the DBP gene.

The P26 gene, a unique feature specific to baculoviruses and consistently positioned, is variably present in some baculoviruses, including instances of a second copy. CaabNPV, for example, two copies of the P26 gene are encoded, with each situated adjacent to P10 and IAP-3, mirroring the presence of two copies of P26 in EranNPV [42].

Except for *Alphabaculovirus urprotei* (UrprNPV), all *Alphabaculoviruses* possess the Chab protein, and typically have two copies [42]. However, in the CaabNPV genome, only a single copy of Chab was evident.

4. Conclusions

In this study, the host was identified as *C. abietis* through morphological examination and COI barcode sequencing, and we first identified and named a baculovirus that was isolated from *C. abietis* larvae collected in Mongolia as *Calliteara abietis* nucleopolyhedrovirus (CaabNPV). SEM analysis revealed the irregularly polyhedra ultrastructure of CaabNPV, consistent with features characteristic of aphabaculoviruses OBs.

Despite the limitation of lacking purified viral strains for detailed analysis of ambiguous regions, high-throughput sequencing and genome assembly provided valuable insights. Based on the estimated pairwise nucleotide distances, CaabNPV has been identified as a novel baculovirus. Phylogenetic analysis placed CaabNPV within the aphabaculoviruses Group II, with OrleNPV identified as its closest relative, showing a high degree of collinearity. Genomic structure analysis unveiled a higher number of ORFs in CaabNPV, some of which are unique compared to other baculoviruses.

In summary, the comprehensive analysis of CaabNPV genome sheds light on its biological features and molecular mechanisms, providing a foundation for research on the diversity, evolution, and host interactions of baculoviruses. Further investigations are warranted to elucidate the ecological roles of CaabNPV and the functional significance of its genomic characteristics.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v16020252/s1>, Table S1: Primer sequences, Table S2: Core genes of 108 sequenced baculovirus genomes, Table S3: CaabNPV genome annotation, Table S4: K2P pairwise distances.

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

Funding: This research was funded by National Key Research and Development Program of China (Grant number 2022YFD1400700), National Natural Science Foundation of China (Grant number 32261143728), The State Key Laboratory of Integrated Management of Pest Insects and Rodents (Grant number IPM2311).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The research data has been successfully uploaded to the open database Genbank to support manuscript submitted.

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 1c.

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

Appendix A

>COI

```
taaagatattggaactctatatttttttgaatttgagcaggaatagtggaaactctttaagattattaattcgagctgaatt-
aggaaatccaggatcattaatcgaaatgatcaaatattatacactattgtgacagctcatgattattataattttttatagttat-
ctattataattggaggattcgaaattgactgtacccttatattaggacacctgatatagctttcctcgaataaataataagtt-
tctgattactcccccttaactttattaatttctagaagaattgtagaaaatggagcaggaacaggatgaacagttacccccac-
ttcatctaattgctcatggaggtagatctgttgattagctatttttctctcatttagctggaattcttcaattttaggagcaaat-
tttattactacaatcattaatgctgattaataattatcttttgatcaataaccattattgtatgagcagttggaattactgctttc-
acttcttctccagttctgctggagctattactatactttaactgatcgtactaaatacatcatttttgaccctgctggaggagg-
agatccaattttatcaacattttttgatttttggtcacc
```

Appendix B

>CaabNPV_contig_189bp
 gatatttgctgttgattatctcgggagacgtaggctccggtttacgaatcgacgggagaacctcacactgacccacc-
 gtttaaattcacgggcgaagggcacatcttgagtgccacactaggctcgtagaggcccgactcgtgatgaatatggctttggtg-
 agtttgatggagtttgagg

>CaabNPV_contig_1171bp (underscored segment representing a 189-bp repetitive sequence).

gatatttgctgttgattatctcgggagacgtaggctccggtttacgaatcgacgggagaacctcacactgacccac-
 cgtttaaattcacgggcgaagggcacatcttgagtgccacactaggctcgtagaggcccgactcgtgatgaatatggctttggtt-
 gccagtttgatggagtttgaggccccgtttttgacgggacctacccattgaatttatcactgctttctacatgcattcgtat-
 ggcttgatcgtatcagagtatccaacacaaaagctactcttttagctttgaaccagggtccccttctcatccacaaatata-
 aaatctcgaagatttatccgcccaggaatgacttgcgaacgaggacatgatgtgttttttacgtgtgagtaattaa-
 aaacaacaccttataataaacggttaaagcaaaattatcttattttgttcttataaaaggatttatgaatacagtagat-
 taattattacataggactcgtataggttggcggcgggcggtgataatggttgatgcaacgattgtggctcgggtgg-
 cgactgttacagtaacttctcaacaaatcaccacgtctctttgttcaatctcattttttgccataactcacacttggaa-
 tcaactcttgatttattcaggggactcactgcgttaggtacattctctctcatcgtagggcataatcattaggatattttg-
 tatagctctcgacataatcttttgcgtgcgaatacactttagcgtatctttatgttgtaaacctgcagttcatgcttgatcca-
 ctattatctggcattacggctatttggcctatcatctacaagttattcattagagataaaatgtgtttctgtcctcggccagtt-
 gagcatctttctgttcaacatgtccaacagtggtgatatttgcgtgtgattatctcgggagacgtaggctccggtttacgaat-
 cgacgggagaacctcacactgacccaccgtttaaattcacgggcgaagggcacatcttgagtgccacactaggctcgtagaggcc-
 cactcgttgatgaatatggctttggtgcccagtttgatggagtttgagg

References

- Ribeiro, B.M.; Dos Santos, E.R.; Trentin, L.B.; da Silva, L.A.; de Melo, F.L.; Kitajima, E.W.; Ardisson-Araújo, D.M.P. A Nymphalid-Infesting Group I Alphabaculovirus Isolated from the Major Passion Fruit Caterpillar Pest *Dione Juno Juno* (Lepidoptera: Nymphalidae). *Viruses* **2019**, *11*, 602. [\[CrossRef\]](#)
- Haase, S.; Sciocco-Cap, A.; Romanowski, V. Baculovirus Insecticides in Latin America: Historical Overview, Current Status and Future Perspectives. *Viruses* **2015**, *7*, 2230–2267. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rohrmann, G.F. *Baculovirus Molecular Biology*, 4th ed.; National Center for Biotechnology Information (US): Bethesda, MD, USA, 2019.
- Ackermann, H.-W.; Smirnoff, W.A. A Morphological Investigation of 23 Baculoviruses. *J. Invertebr. Pathol.* **1983**, *41*, 269–280. [\[CrossRef\]](#)
- Blissard, G.W.; Theilmann, D.A. Baculovirus Entry and Egress from Insect Cells. *Annu. Rev. Virol.* **2018**, *5*, 113–139. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nagamine, T.; Sako, Y. A Role for the Anti-Viral Host Defense Mechanism in the Phylogenetic Divergence in Baculovirus Evolution. *PLoS ONE* **2016**, *11*, e0156394. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hill, T.; Thickless, R.L. Baculovirus Molecular Evolution via Gene Turnover and Recurrent Positive Selection of Key Genes. *J. Virol.* **2017**, *91*, e01319-17. [\[CrossRef\]](#) [\[PubMed\]](#)
- Thézé, J.; Lopez-Vaamonde, C.; Cory, J.S.; Herniou, E.A. Biodiversity, Evolution and Ecological Specialization of Baculoviruses: A Treasure Trove for Future Applied Research. *Viruses* **2018**, *10*, 366. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gninenko, Y.I.; Serdyukov, G.V. The Coniferous Tussock Moth (*Calliteara abietis*, Lepidoptera, Lymantriidae) in Buryatia. *Entomol. Rev.* **2010**, *90*, 999–1002. [\[CrossRef\]](#)
- Shu, R.; Meng, Q.; Miao, L.; Liang, H.; Chen, J.; Xu, Y.; Cheng, L.; Jin, W.; Qin, Q.; Zhang, H. Genome Analysis of a Novel Clade b *Betabaculovirus* Isolated from the Legume Pest *Matsumuraes phaseoli* (Lepidoptera: Tortricidae). *Viruses* **2020**, *12*, 1068. [\[CrossRef\]](#)
- Stewart, D.; Zahiri, R.; Djoumad, A.; Freschi, L.; Lamarche, J.; Holden, D.; Cervantes, S.; Ojeda, D.I.; Potvin, A.; Nisole, A.; et al. A Multi-Species TaqMan PCR Assay for the Identification of Asian Gypsy Moths (*Lymantria* spp.) and Other Invasive Lymantriines of Biosecurity Concern to North America. *PLoS ONE* **2016**, *11*, e0160878. [\[CrossRef\]](#)
- Zhang, H.; Tong, Y.; Liu, H.; Guo, L.; Jin, W.; Li, X.; Meng, Q.; Yu, X.; Fang, F.; Qin, Q.; et al. Establishment of a Rapid Detection Method for *Plutella xylostella* Granulovirus Based on qPCR. *Heliyon* **2023**, *9*, e15170. [\[CrossRef\]](#)
- Denman, A.M. Molecular Cloning: A Laboratory Manual. *Immunology* **1983**, *49*, 411.

15. Solovyev, V.; Kosarev, P.; Seledsov, I.; Vorobyev, D. Automatic Annotation of Eukaryotic Genes, Pseudogenes and Promoters. *Genome Biol.* **2006**, *7*, S10. [[CrossRef](#)]
16. van Oers, M.M.; Vlak, J.M. Baculovirus genomics. *Curr. Drug Targets* **2007**, *8*, 1051–1068. [[CrossRef](#)] [[PubMed](#)]
17. Hu, Z.H.; Arif, B.M.; Jin, F.; Martens, J.W.; 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 Pt 11*, 2841–2851. [[CrossRef](#)] [[PubMed](#)]
18. Madeira, F.; Pearce, M.; Tivey, A.R.N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and Sequence Analysis Tools Services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022**, *50*, W276–W279. [[CrossRef](#)] [[PubMed](#)]
19. Benson, G. Tandem Repeats Finder: A Program to Analyze DNA Sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [[CrossRef](#)] [[PubMed](#)]
20. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276–277. [[CrossRef](#)] [[PubMed](#)]
21. 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](#)]
22. Guarino, L.A.; Smith, M. Regulation of Delayed-Early Gene Transcription by Dual TATA Boxes. *J. Virol.* **1992**, *66*, 3733–3739. [[CrossRef](#)]
23. 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](#)]
24. Wormleaton, S.; Kuzio, J.; Winstanley, D. The Complete Sequence of the Adoxophyes Orana Granulovirus Genome. *Virology* **2003**, *311*, 350–365. [[CrossRef](#)]
25. Morris, T.D.; Miller, L.K. Mutational Analysis of a Baculovirus Major Late Promoter. *Gene* **1994**, *140*, 147–153. [[CrossRef](#)]
26. 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](#)]
27. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics. *Genome Biol.* **2019**, *20*, 238. [[CrossRef](#)]
28. Kozlov, A.M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference. *Bioinformatics* **2019**, *35*, 4453–4455. [[CrossRef](#)]
29. Lanfear, R.; Frandsen, P.B.; Wright, A.M.; Senfeld, T.; Calcott, B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol. Biol. Evol.* **2017**, *34*, 772–773. [[CrossRef](#)] [[PubMed](#)]
30. 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](#)]
31. 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](#)]
32. Friesen, P.D.; Nissen, M.S. Gene Organization and Transcription of TED, a Lepidopteran Retrotransposon Integrated within the Baculovirus Genome. *Mol. Cell Biol.* **1990**, *10*, 3067–3077. [[CrossRef](#)] [[PubMed](#)]
33. Beames, B.; Summers, M.D. Sequence Comparison of Cellular and Viral Copies of Host Cell DNA Insertions Found in Autographa Californica Nuclear Polyhedrosis Virus. *Virology* **1990**, *174*, 354–363. [[CrossRef](#)] [[PubMed](#)]
34. Clavijo, G.; Williams, T.; Muñoz, D.; Caballero, P.; López-Ferber, M. Mixed Genotype Transmission Bodies and Virions Contribute to the Maintenance of Diversity in an Insect Virus. *Proc. R. Soc. B Biol. Sci.* **2010**, *277*, 943–951. [[CrossRef](#)]
35. Miele, S.A.B.; Garavaglia, M.J.; Belaich, M.N.; Ghiringhelli, P.D. Baculovirus: Molecular Insights on Their Diversity and Conservation. *Int. J. Evol. Biol.* **2011**, *2011*, 379424. [[CrossRef](#)] [[PubMed](#)]
36. Thumbi, D.K.; Eveleigh, R.J.M.; Lucarotti, C.J.; Lapointe, R.; Graham, R.I.; Pavlik, L.; Lauzon, H.A.M.; Arif, B.M. Complete Sequence, Analysis and Organization of the Orgyia Leucostigma Nucleopolyhedrovirus Genome. *Viruses* **2011**, *3*, 2301–2327. [[CrossRef](#)] [[PubMed](#)]
37. 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](#)]
38. Ardisson-Araújo, D.M.P.; da Silva, A.M.R.; Melo, F.L.; Dos Santos, E.R.; Sosa-Gómez, D.R.; Ribeiro, B.M. A Novel Betabaculovirus Isolated from the Monocot Pest *Mocis latipes* (Lepidoptera: Noctuidae) and the Evolution of Multiple-Copy Genes. *Viruses* **2018**, *10*, 134. [[CrossRef](#)] [[PubMed](#)]
39. Guarino, L.A.; Gonzalez, M.A.; Summers, M.D. Complete Sequence and Enhancer Function of the Homologous DNA Regions of Autographa Californica Nuclear Polyhedrosis Virus. *J. Virol.* **1986**, *60*, 224–229. [[CrossRef](#)]
40. 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](#)]
41. Kang, W.K.; Imai, N.; Suzuki, M.; Iwanaga, M.; Matsumoto, S.; Zemskov, E.A. Interaction of Bombyx Mori Nucleopolyhedrovirus BRO-A and Host Cell Protein Laminin. *Arch. Virol.* **2003**, *148*, 99–113. [[CrossRef](#)]
42. Liu, L.; Zhang, Z.; Liu, C.; Qu, L.; Wang, D. Genome Analysis of an Alphabaculovirus Isolated from the Larch Looper, *Erannis ankeraria*. *Viruses* **2021**, *14*, 34. [[CrossRef](#)] [[PubMed](#)]

43. Mikhailov, V.S.; Vanarsdall, A.L.; Rohrmann, G.F. Isolation and Characterization of the DNA-Binding Protein (DBP) of the Autographa Californica Multiple Nucleopolyhedrovirus. *Virology* **2008**, *370*, 415–429. [[CrossRef](#)] [[PubMed](#)]
44. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.