Virus Genes (2019) 55:104–116
https://doi.org/10.1007/s11262-018-1615-7

The complete genome sequence of a second alphabaculovirus from the true armyworm, *Mythimna unipuncta*: implications for baculovirus phylogeny and host specificity

Robert L. Harrison[1] · Joseph D. Mowery[2] · Gary R. Bauchan[2] · David A. Theilmann[3] · Martin A. Erlandson[4]

Received: 30 August 2018 / Accepted: 9 November 2018 / Published online: 14 November 2018

Abstract

The Mythimna unipuncta nucleopolyhedrovirus isolate KY310 (MyunNPV-KY310) is an alphabaculovirus isolated from a true armyworm (*Mythimna unipuncta*) population in Kentucky, USA. Occlusion bodies of this virus were examined by electron microscopy and the genome sequence was determined by 454 pyrosequencing. MyunNPV-KY310 occlusion bodies consisted of irregular polyhedra measuring 0.8–1.8 µm in diameter and containing multiple virions, with one to six nucleocapsids per virion. The genome sequence was determined to be 156,647 bp with a nucleotide distribution of 43.9% G+C. 152 ORFs and six homologous repeat (hr) regions were annotated for the sequence, including the 38 core genes of family *Baculoviridae* and an additional group of 26 conserved alphabaculovirus genes. BLAST queries and phylogenetic inference confirmed that MyunNPV-KY310 is most closely related to the alphabaculovirus Leucania separata nucleopolyhedrovirus isolate AH1, which infects *Mythimna separata*. In contrast, MyunNPV-KY310 did not exhibit a close relationship with Mythimna unipuncta nucleopolyhedrovirus isolate #7, an alphabaculovirus from the same host species. MyunNPV-KY310 lacks the *gp64* envelope glycoprotein, which is a characteristic of group II alphabaculoviruses. However, this virus and five other alphabaculoviruses lacking *gp64* are placed outside the group I and group II clades in core gene phylogenies, further demonstrating that viruses of genus *Alphabaculovirus* do not occur in two monophyletic clades. Potential instances of MyunNPV-KY310 ORFs arising by horizontal transfer were detected. Although there are now genome sequences of four different baculoviruses from *M. unipuncta*, comparison of their genome sequences provides little insight into the genetic basis for their host specificity.

Keywords Baculovirus · Alphabaculovirus · Nucleopolyhedrovirus · *Mythimna unipuncta* · True armyworm · MyunNPV

Introduction

Baculoviruses are viruses of insects that possess large, double-stranded circular DNA genomes [1]. These viruses produce enveloped, rod-shaped virions that are embedded in proteinaceous occlusion bodies (OBs). Species of baculoviruses are organized into four genera within a single family, *Baculoviridae* [1]. Two of these genera, *Alphabaculovirus* and *Betabaculovirus*, contain 65 of the 68 species currently classified for *Baculoviridae*. Viruses of both genera infect larvae of the insect order Lepidoptera. Decades of research with isolates of these two genera have uncovered extensive information about the genetics, pathology, and life cycle of baculoviruses [2, 3] and led to applications as recombinant protein expression vectors and safe, ecologically friendly biopesticides [4, 5].

Edited by A. Lorena Passarelli.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11262-018-1615-7) contains supplementary material, which is available to authorized users.

✉ Robert L. Harrison
Robert.L.Harrison@ars.usda.gov

[1] Invasive Insect Biocontrol and Behavior Laboratory, Beltsville Agricultural Research Center, USDA Agricultural Research Service, Beltsville, MD 20705, USA

[2] Electron and Confocal Microscopy Unit, Beltsville Agricultural Research Center, USDA Agricultural Research Service, Beltsville, MD 20705, USA

[3] Summerland Research and Development Centre, Agriculture and Agri-Food Canada, Summerland, BC V0H 1Z0, Canada

[4] Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, SK S7N 0X2, Canada

In this study, the complete genome sequence of the alphabaculovirus Mythimna unipuncta nucleopolyhedrovirus KY310 (MyunNPV-KY310) is detailed. This virus was originally isolated from larval cadavers of the true armyworm, *Mythimna unipuncta*, collected in a Lexington, KY (USA) pasture during an outbreak [6]. Analysis of partial nucleotide sequences of the *lef-8, lef-9*, and *polyhedrin* (*polh*) genes of MyunNPV-KY310 suggested that it represents one of four different and distinct groups of baculoviruses isolated from true armyworm [6]. Genomes from viruses of three of these groups have been previously sequenced and analyzed, including the betabaculoviruses Pseudaletia unipuncta granulovirus Hawaiian (PsunGV-H, GenBank accession no. EU678671; species *Mythimna unipuncta granulovirus A*) and Mythimna unipuncta granulovirus #8 [7] (MyunGV#8; species *Mythimna unipuncta granulovirus B*) and the alphabaculovirus Mythimna unipuncta nucleopolyhedrovirus #7 [8] (MyunNPV#7; unclassified as of this writing). Examination of the MyunNPV-KY310 OBs and genome sequence in this study confirmed that it likely represents a fourth distinct species of baculovirus to be isolated from *M. unipuncta*. Comparison of the MyunNPV-KY310 genome sequence with other *Mythimna* sp. baculovirus genome sequences highlights the divergence among alphabaculoviruses that exists even among viruses from the same host, and underscores the difficulties with attempting to define the genetic basis for baculovirus host range.

Materials and methods

Virus

MyunNPV-KY310 OBs were isolated from virus-killed *M. unipuncta* cadavers collected from a pasture in Lexington, KY during March, 2010. The homogenization of cadavers and purification of OBs by low-speed centrifugation were carried out using previously described procedures [8, 9].

Electron microscopy

Methods for cryogenic scanning electron microscopy (Cryo-SEM) and transmission electron microscopy (TEM) of MyunNPV-KY310 OBs were as previously described [8].

Viral DNA isolation and sequencing

Viral genomic DNA was extracted from the occlusion-derived virus of 1.5×10^{10} OBs of MyunNPV-KY310 using previously described procedures [10]. Determination of the genome sequence was carried out at the National Research Council, Plant Biotechnology Institute (Saskatoon, Saskatchewan, Canada) using Roche 454 FLX-titanium pyrosequencing technology. Sequence reads initially were assembled using CLC-Genomics Workbench 6.0.2, with additional assembly and read mapping carried out with Lasergene SeqMan Pro version 11. Regions of the genome with repeats or unusual features were amplified by PCR and sequenced by Sanger dideoxy sequencing.

The MyunNPV-KY310 genome sequence generated during this study has been deposited in GenBank with the accession number MH124167.

Genome sequence analysis

Open reading frames (ORFs) that are homologs of previously annotated baculovirus genes were identified by BLASTp queries of conceptual amino acid sequences. Other ORFs were annotated if they were ≥ 50 codons in size, did not overlap larger ORFs by > 75 bp, and were predicted to encode proteins by both the fgenesV (http://linux1.softberry.com/berry.phtml) and ZCURVE_V [11] programs. Amino acid sequences from such ORFs were also used in HHpred queries [12] to identify conserved domains.

Repeated sequences in homologous repeat (*hr*) and unique repeat (*ur*) regions were identified using LaserGene Genequest v. 15 (DNASTAR).

Sequence comparisons and phylogenies

The order of homologous genes between MyunNPV-KY310 and selected alphabaculoviruses was determined using gene parity plots [13]. Synteny between the MyunNPV-KY310 and LeseNPV-AH1 genomes was further evaluated by alignment of the genome sequences using Mauve [14] as implemented in LaserGene MegAlign Pro 15, with a seed weight of 15.

To infer a core gene phylogeny, core gene amino acid sequences from the baculovirus genomes listed in Supplementary Table 1 were aligned using the MegaAlign Pro 15 implementation of MUSCLE [15]. Alignments were concatenated with BioEdit [16]. Phylogeny was inferred by (1) the minimum evolution (ME) method in MEGA7 [17], using the JTT substitution matrix with a gamma shape parameter of 0.81 and 500 bootstrap iterations, and (2) the maximum likelihood (ML) method in RAxML [18] using the LG matrix. Alignments of amino acid sequences encoded by selected *he65* and GIY-YIG nuclease (*ac79*) homologs, and phylogenies based on those alignments, were carried out with the same methods. The GIY-YIG nuclease ME and ML trees were inferred with MEGA7 and JTT or LG matrices, respectively, with a gamma shape parameter of 2.9. Both ME and ML *he65* trees were inferred using MEGA7 and the JTT matrix, with a gamma shape parameter of 1.4.

The consensus sequences of unit repeats in MyunNPV-KY310 and LeseNPV-AH1 "unique" and homologous repeat

regions were determined from MUSCLE alignments of individual unit repeat sequences.

Pairwise nucleotide distances for partial *lef-8, lef-9*, and *polh* sequences were estimated between MyunNPV-KY310 and other alphabaculoviruses using MEGA7 and the Kimura-2-parameter substitution model as previously described [19], to address whether MyunNPV-KY310 met the criteria described by Jehle et al. [19] for belonging to a new baculovirus species.

Results

Occlusion bodies (OBs) and virions

MyunNPV-KY310 OBs exhibited an irregular polyhedral shape typical of alphabaculovirus OBs (Fig. 1a, b). The OBs in SEM images measured from 0.8 to 1.8 µm in diameter, a range consistent with sizes reported for OBs of other alphabaculoviruses [20].

Several enveloped virions were observed within the matrix of the OBs (Fig. 1c, d). Each virion contained from one to six nucleocapsids, with most virions appearing to contain two or three nucleocapsids. Nucleocapsids measured approximately 300 nm in length, consistent with reported lengths for other alphabaculovirus nucleocapsids [20].

Basic properties of the genome sequence

Assembly of MyunNPV-KY310 sequencing reads yielded a circular genome sequence of 156,647 bp with a coverage of 433× (Fig. 2). The genome exhibited a nucleotide distribution of 43.9% G+C. The genome sequence was linearized at the start-codon adenine of the polyhedrin (*polh*) ORF and deposited in GenBank (accession number MH124167).

A total of 152 ORFs and six *homologous repeat* (*hr*) regions were identified and annotated in the MyunNPV-KY310 genome (Fig. 2, Supplementary Table 2). Annotated ORFs included those for the 38 core genes that have been found in all baculovirus genomes to date [21, 22]. The *hr* regions of MyunNPV-KY310 consisted of 12-bp and 14-bp palindromes very similar to those reported for the *hr* regions of the Leucania separata nucleopolyhedrovirus genome (isolate LeseNPV-AH1; [23]). The consensus

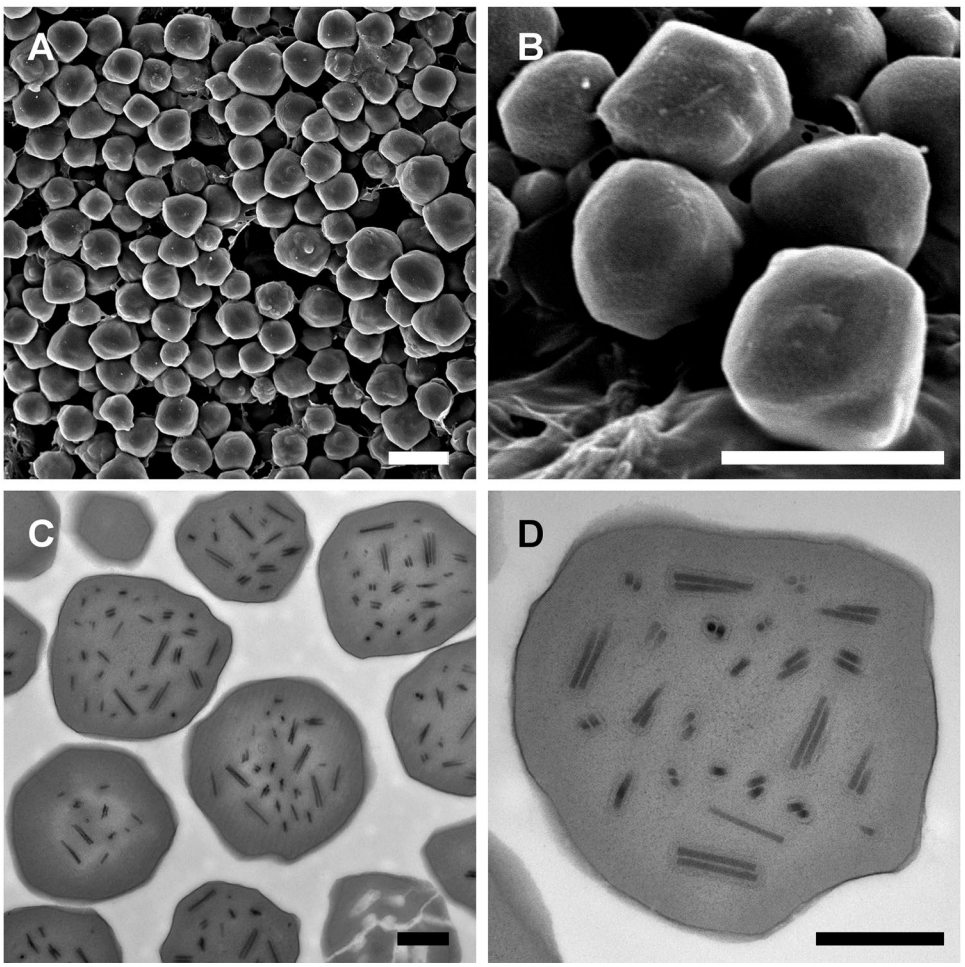

Fig. 1 Occlusion bodies of *Mythimna unipuncta* nucleopolyhedrovirus isolate KY310 (MyunNPV-KY310). **a, b** Scanning electron micrographs of occlusion bodies. **c**, **d** Transmission electron micrographs of ultrathin sections through occlusion bodies. Scale bars: 2 µm (**a**, **b**), 500 nm (**c**, **d**)

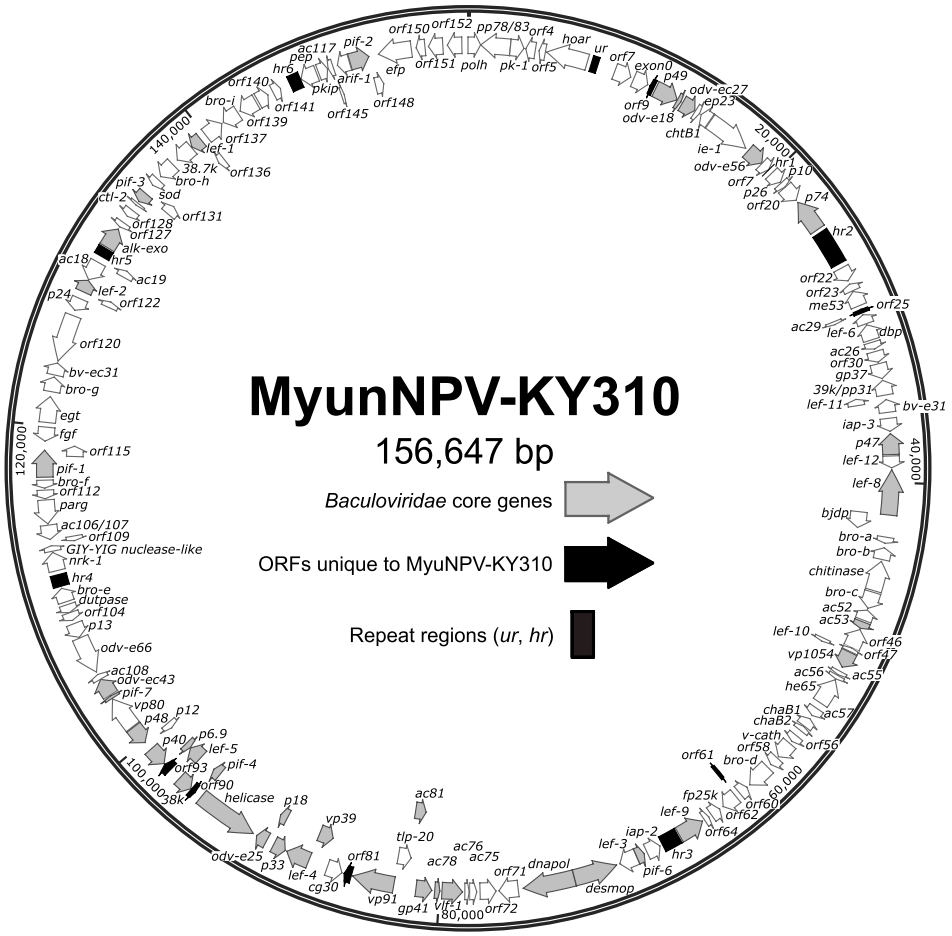

Fig. 2 Map of the ORFs and other features of the MyunNPV-KY310 genome. The position and orientation of ORFs on the circular genome are represented by arrows, with gray arrows representing core genes of family *Baculoviridae* and black arrows representing ORFs unique to MyunNPV-KY310. Each ORF is designated by either its number in the genome annotation or the abbreviation of its homolog. The presence and locations of regions consisting of sequence repeats are also indicated

sequence of the 12-bp palindromes, referred to by Xiao et al. [23] as P-I repeats, was identical to the reported sequence of the LeseNPV-AH1 sequence (5′-CGAACCGGTTCG-3′). The 14-bp (P-II) repeat consensus differs at two nucleotide positions between the two viruses: in the LeseNPV-AH1 genome, the P-II consensus is 5′-CGTTGGGCCCAACG-3′, while the MyunNPV-KY310 P-II consensus is CGTTGG ATCCAACG-3′. A "unique repeat" (*ur*) region, reported to occur between the ORF6 (*hoar*) and ORF7 in the LeseNPV-AH1 sequence, also was detected in the same location in MyunNPV-KY310. Both *ur* regions consist of six or seven similar, adenine-rich direct repeats (Fig. 3).

Relationships with other baculoviruses

Phylogenetic relationships among MyunNPV-KY310 and other baculoviruses were inferred from concatenated alignments of the 38 baculovirus core gene amino acid sequences. Alphabaculoviruses formed a single monophyletic group, and group I alphabaculoviruses [24, 25] formed a single monophyletic clade within the alphabaculovirus group (Fig. 4). The remaining alphabaculoviruses did not form a monophyletic clade, but were paraphyletic in both the ME and ML trees.

MyunNPV-KY310 was placed in a clade with LeseNPV-AH1, which is consistent with results of a previously published phylogeny based on concatenated *lef-8*/*lef-9*/*polh* alignments [6] and BLASTp queries showing that MyunNPV-KY310 is most closely related to LeseNPV-AH1 (Supplementary Table 2). LeseNPV-AH1 is a virus of *Mythimna* (*Leucania*) *separata*, a species that is closely related to *M. unipuncta*. MyunNPV-KY310 and LeseNPV-AH1 were grouped into a larger clade with Spodoptera littoralis nucleopolyhedrovirus (SpliNPV-AN1956 [26]) and Spodoptera litura nucleopolyhedrovirus (SpltNPV-G2 [27]). Notably, Mythimna unipuncta nucleopolyhedrovirus #7, a virus from the same host species [8], was placed in a different part of the alphabaculovirus phylogeny (Fig. 4).

Other Genbank entries of nucleotide sequences from a *Leucania separata* nucleopolyhedrovirus were submitted prior to the LeseNPV-AH1 genome sequence, including one 1,446 bp sequence [28] and a 5,423 bp sequence [29]. However, BLAST queries with nucleotide and amino acid sequences from these GenBank submissions indicated that they come from viruses that are variants of Mamestra

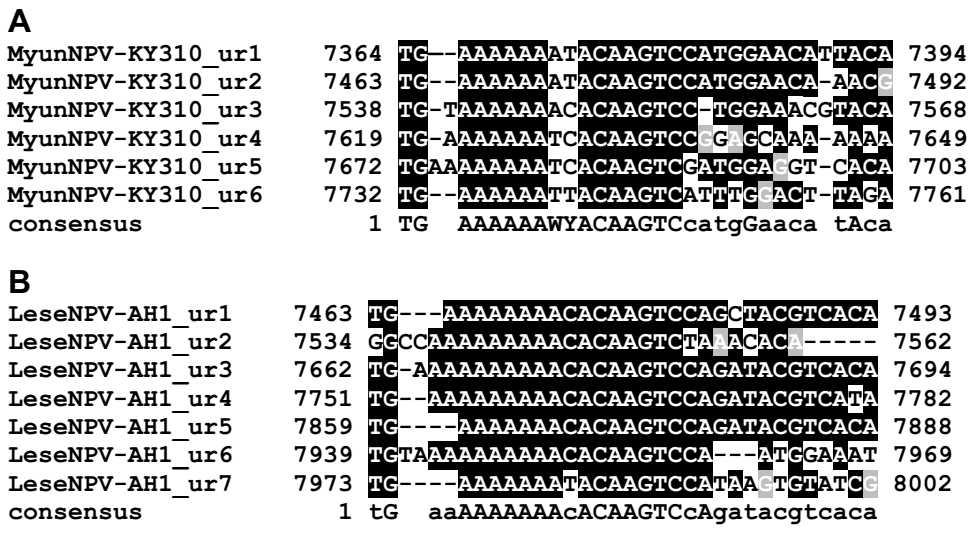

Fig. 3 Sequences of unit repeats of the "unique repeat" (*ur*) regions of **a** MyunNPV-KY310, and **b** LeseNPV-AH1. Identical nucleotides occupying ≥ 50% of aligned positions are shaded in black, and nucleotides of the same class as conserved nucleotides (containing either a purine or pyrimidine base) are shaded in gray. Nucleotides in the consensus sequence are denoted by uppercase letters for positions in the alignment with 100% identical residues, and lowercase letters for positions in the alignment with ≥ 50%, but < 100%, identical residues. IUPAC base codes are as follows: W = A or T; Y = C or T

configurata nucleopolyhedrovirus B (MacoNPV-B) and not from isolate LeseNPV-AH1 (data not shown).

Gene parity plots with selected alphabaculoviruses revealed a significant degree of collinearity between MyunNPV-KY310 and the genomes of LeseNPV-AH1 and SpltNPV-G2, and noticeably less collinearity with HearNPV-G4 and MyunNPV #7 (Fig. 5). Examination of the plots revealed a region of the MyunNPV-KY310 genome encompassing the ORFs between ORF114 (*pif-1*) and ORF133 (*bro-h*) which was inverted with respect to the corresponding regions in LeseNPV-AH1 and SpltNPV-G2 (Fig. 5, dashed-line box). Mauve alignment of the MyunNPV-KY310 and LeseNPV-AH1 nucleotide sequences confirmed that a locally collinear block (LCB) extending from nt 116,955 to 137,679 in the MyunNPV-KY310 genome sequence, containing ORFs 114–133, was present in the opposite orientation in the LeseNPV-AH1 sequence (data not shown).

Nucleotide distances between partial *lef-8, lef-9*, and *polh* sequences of MyunNPV-KY310 and LeseNPV-AH1 were estimated at 0.397, 0.232, and 0.294 substitutions/site, respectively, indicating that these two alphabaculoviruses belong to different species [19]. Larger values were seen for pairwise distances with other isolates of currently recognized alphabaculovirus species, suggesting that MyunNPV-KY310 represents an undescribed species of genus *Alphabaculovirus*. A recent application of a phylogenetic clustering approach to baculovirus species delimitation [30] also indicated that MyunNPV-KY310 represented a species separate from *Leucania separata nucleopolyhedrovirus*.

Gene content

ORFs annotated for the MyunNPV-KY310 genome sequence included homologs for 26 genes common to all alphabaculoviruses [21]. No homolog for ubiquitin (*ac35*) was found in MyunNPV-KY310, a feature previously reported for the LeseNPV-AH1 genome sequence [23]. Nine members of the *baculovirus repeated ORF* (*bro*) multigene family [31] were found in the MyunNPV-KY310 sequence. Eight of the nine *bro* family members in the MyunNPV-KY310 sequence also were found in identical locations within the LeseNPV-AH1 genome (Supplementary Table 2). BLAST queries indicate that 14 MyunNPV-KY310 ORFs (4, 5, 7, 23, 64, 71, 104, 109, 131, 145, 148, 150, 151, and 152) were unique to the MyunNPV-KY310 and LeseNPV-AH1 genomes, and six ORFs (9, 25, 61, 81, 90, and 93) occurred solely in the MyunNPV-KY310 sequence. Of these 20 ORFs, database queries with 19 failed to detect conserved domains or sequence similarity with other genes. An HHpred query with ORF150 (LeseNPV-AH1 ORF167) revealed the presence of domain DUF3587 (probability = 100%), which is found in ichnovirus repeat element (*rep*) proteins [32, 33].

Horizontal gene transfer has been shown to account for some of the gene content of baculoviruses [34]. BLASTp queries produced evidence of horizontal gene transfer involving some MyunNPV-KY310 ORFs. The top database search matches with the adjacent ORFs 128 and 129 (*ctl-2*) were with ORFs from clade *a* betabaculoviruses (Supplementary Table 2), and the only match obtained with a query using ORF141 was with *Agrotis segetum* nucleopolyhedrovirus A (AgseNPV-A; [35]) ORF61.

MyunNPV-KY310 ORF52 is a homologue of *he65* (*ac105*), a gene expressed early during infection and found in a subset of alphabaculovirus and betabaculovirus genomes [36]. BLASTp analysis of the MyunNPV-KY310 *he65* homolog yielded top matches with the *he65* gene in Xestia c-nigrum granulovirus (XecnGV) and related betabaculoviruses (Supplementary Table 2). Phylogenetic inference with other baculovirus and entomopoxvirus homologs of *he65* placed the MyunNPV-KY310 and LeseNPV-AH1 *he65*

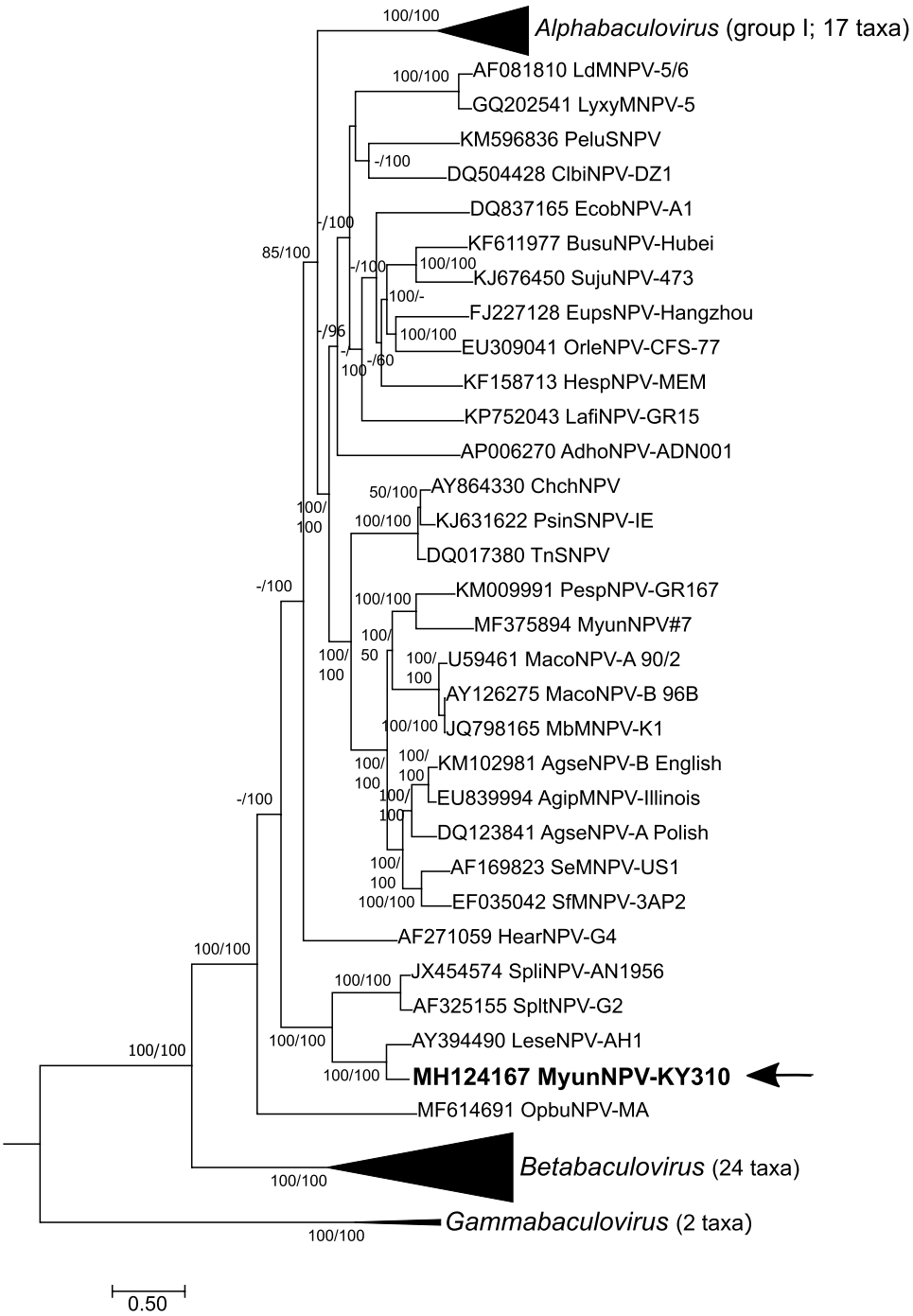

Fig. 4 Core gene phylogeny showing the relationships among MyunNPV-KY310, representative isolates of baculovirus species, and other unclassified isolates. The phylogenetic tree was inferred from the concatenated alignments of 38 baculovirus core gene amino acid sequences using the ML method, with the deltabaculovirus Culex nigripalpus nucleopolyhedrovirus (CuniNPV) used as an outgroup (not shown). Nodes for interior branches show bootstrap values of > 50% where they occur for both ME and ML analyses of the concatenated alignments, with the ME bootstrap value listed before the slash and the ML value listed after the slash (ME/ML). Dashes indicate branches with < 50% bootstrap in the ME or ML tree. Branches for the group I viruses of genus *Alphabaculovirus* and the viruses of genera *Betabaculovirus* and *Gammabaculovirus* are compressed. Terminal branches shown correspond to alphabaculoviruses that do not possess a *gp64* gene (group II). Genome accession numbers and isolate abbreviations are displayed for these viruses. All virus taxa used in the analysis are listed in Supplementary Table 1

amino acid sequences in a well-supported group with other clade *a* betabaculovirus HE65 sequences (Fig. 6). This group in turn was part of a larger clade that included other group II alphabaculovirus HE65 sequences. The HE65 sequences of group I alphabaculoviruses also formed a well-supported clade, but branches for two betabaculovirus sequences also occurred in this group, suggesting horizontal transfer among the group I alphabaculoviruses and betabaculoviruses. Branches for other HE65 sequences from entomopoxviruses and from other individual alphabaculovirus and betabaculovirus sequences were poorly supported.

A BLASTp search with MyunNPV-KY310 ORF108 resulted in several matches to GIY-YIG nucleases from bacteria (Supplementary Table 2). The GIY-YIG superfamily includes nucleases that occur in a broad range of organisms and are distinguished by an approximately 70- to 100-amino acid domain containing the motifs GIY and YIG, followed by conserved arginine and glutamate residues [37]. GIY-YIG

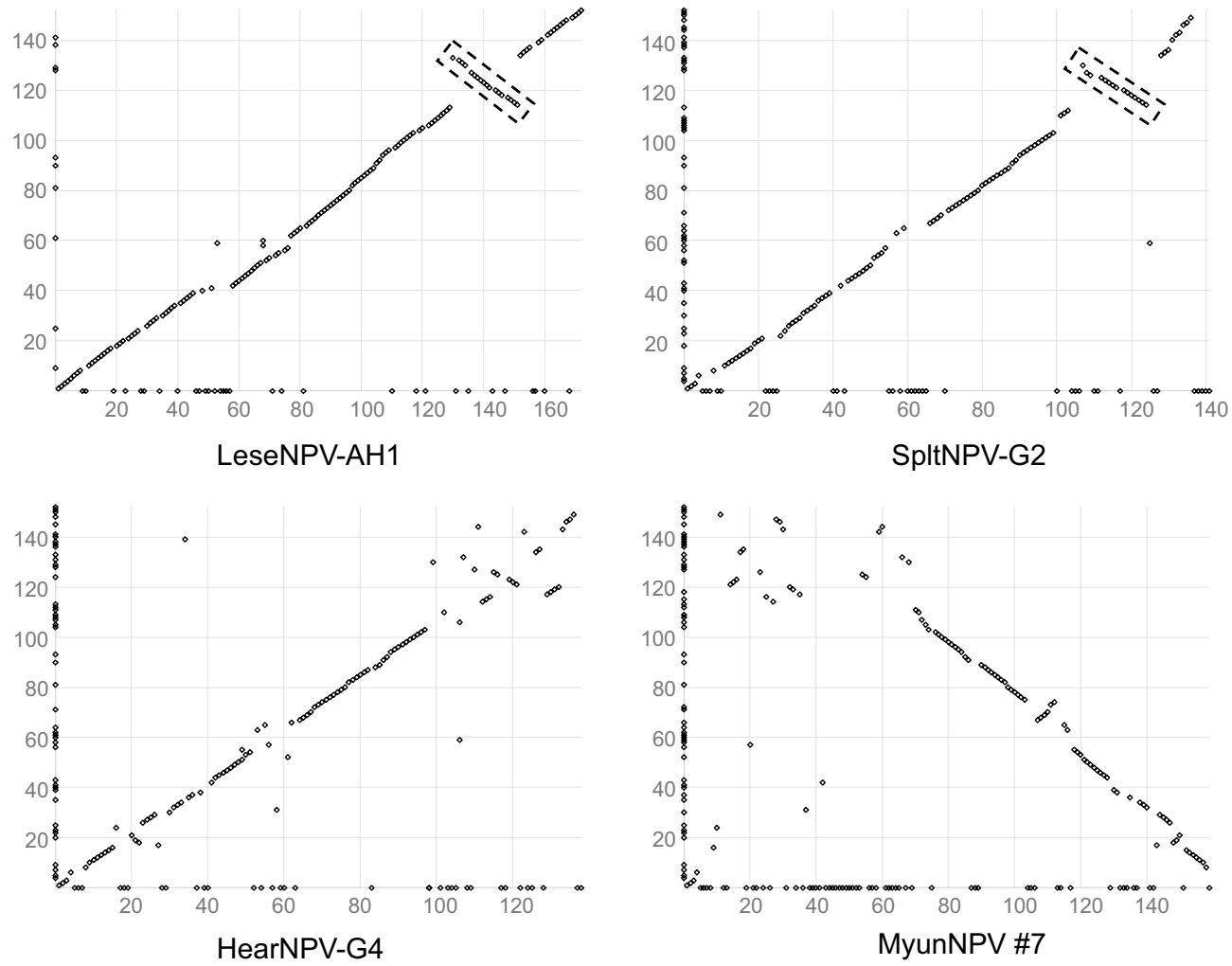

Fig. 5 The ORF content and order of the MyunNPV-KY310 genome compared with that of LeseNPV-AH1, Helicoverpa armigera nucleopolyhedrovirus G4 (HearNPV-G4), Spodoptera littoralis nucleopolyhedrovirus AN1956 (SpliNPV-AN1956), and MyunNPV#7 by gene parity plots. Each point in a plot represents an ORF. ORFs present in only one of the genomes being compared appear on the axis corresponding to the virus in which they are present (*y*-axis for MyunNPV-KY310, *x*-axis for the other viruses). A large group of contiguous ORFs that are present in an inverted order in MyunNPV-KY310 compared to LeseNPV-AH1 and SpliNPV-AN1956 are indicated by a dashed-line box

nuclease homologs have been previously identified in alpha- and beta-baculoviruses [38]. However, no baculovirus GIY-YIG homologs appeared among BLAST results with ORF108 other than the homologous ORF121 encoded by LeseNPV-AH1. Phylogenies inferred from an alignment of ORF108, other baculovirus and ascovirus Ac79 homologs, and selected bacterial GIY-YIG nuclease homologs placed the group I alphabaculovirus Ac79 homologs, the betabaculovirus clade *a* homologs, and the bacterial GIY-YIG family members each into their own groups (Fig. 7). The MyunNPV-KY310 and LeseNPV-AH1 sequences were not placed with the bacterial sequences, but instead occurred along with *Orgyia leucostigma* nucleopolyhedrovirus (OrleNPV-CFS-77) ORF97 in a moderately well-supported clade with 62%/72% ME/ML bootstrap support. Alignment of ORF108 with a subset of these sequences revealed that ORF108 possessed copies of the GIY and YIG motifs and the conserved arginine (Fig. 8). However, ORF108 and OrleNPV-CFS-77 ORF97 encoded a lysine in place of the conserved glutamate.

Discussion

Until recently, core gene phylogenies of baculoviruses placed the alphabaculoviruses into two monophyletic clades, referred to as group I and group II [25] (for recent examples, see [39, 40]). Publications on recently determined

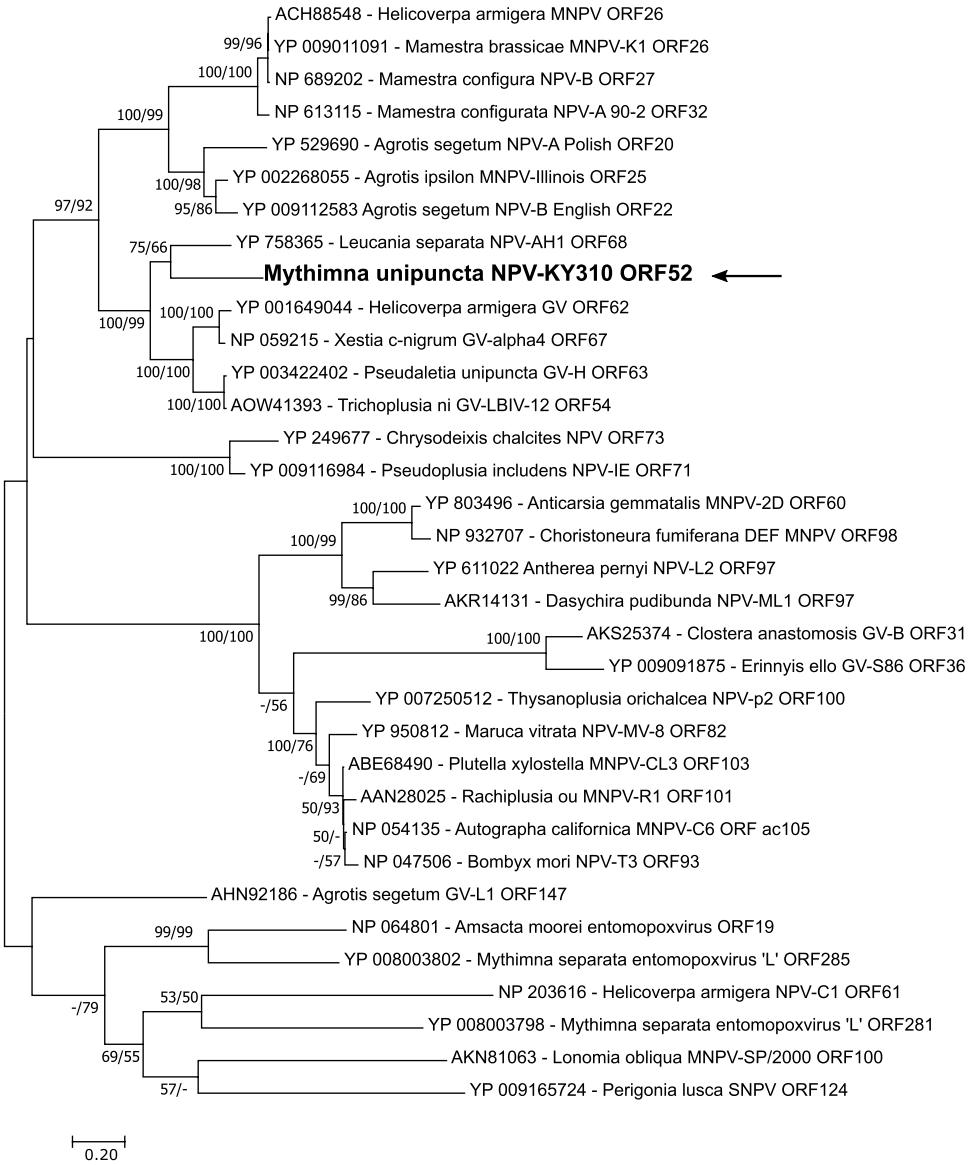

Fig. 6 Phylogeny of baculovirus and entomopoxvirus HE65 homologs. ML and ME phylograms were inferred from the alignment of *he65*-encoded amino acid sequences of alphabaculoviruses (NPVs), betabaculoviruses (GVs), and two entomopoxviruses, with ORF and GenBank accession numbers displayed for each sequence. The ML tree is shown with bootstrap values of ≥ 50% for the ME and ML trees (ME/ML) where they occur, as described for Fig. 4. The MyunNPV-KY310 HE65 is indicated with an arrow

alphabaculovirus genome sequences have reported phylogenies in which not all alphabaculoviruses occur in two monophyletic groups. Five viruses—LeseNPV-AH1, Operophtera brumata nucleopolyhedrovirus-MA (OpbuNPV-MA), Helicoverpa armigera nucleopolyhedrovirus-G4 (HearNPV-G4), SpltNPV-G2, and SpliNPV-AN1956—have been placed on branches outside of the clade containing the group I and other group II alphabaculoviruses [8, 30, 41, 42]. Our analysis has shown that MyunNPV-KY310 is also a member of this group of viruses (Fig. 4). Harrison et al. [8] reported that these viruses could be placed back on the same basal node as the other group II alphabaculoviruses in an ME phylogeny based on MUSCLE alignments from which sequences of the Urbanus proteus nucleopolyhedrovirus (UrprNPV) had been excluded. However, with MyunNPV-KY310 sequences present in the core gene amino acid alignments, exclusion of UrprNPV sequences did not result in a reunion of all group II viruses into a single monophyletic clade. This suggests that the addition of more taxa to alphabaculovirus core gene trees will likely reinforce the paraphyly of the group II alphabaculoviruses.

As is typical for baculovirus genomes, some ORFs were identified in the MyunNPV-KY310 genome whose presence may be due to horizontal transfer. Of particular interest were ORFs 52 (*he65*) and 108 (GIY-YIG nuclease). The ancestral ORF 52 sequence apparently was either obtained from clade *a* betabaculoviruses or was a source for *he65* in the genomes of these viruses (Fig. 6). The *he65* gene product was initially identified as one of a set of genes that induced nuclear localization of actin in transient expression assays [43]. Subsequent studies with *he65* knockout mutants found that *he65* was not required for actin nuclear localization [44], with a

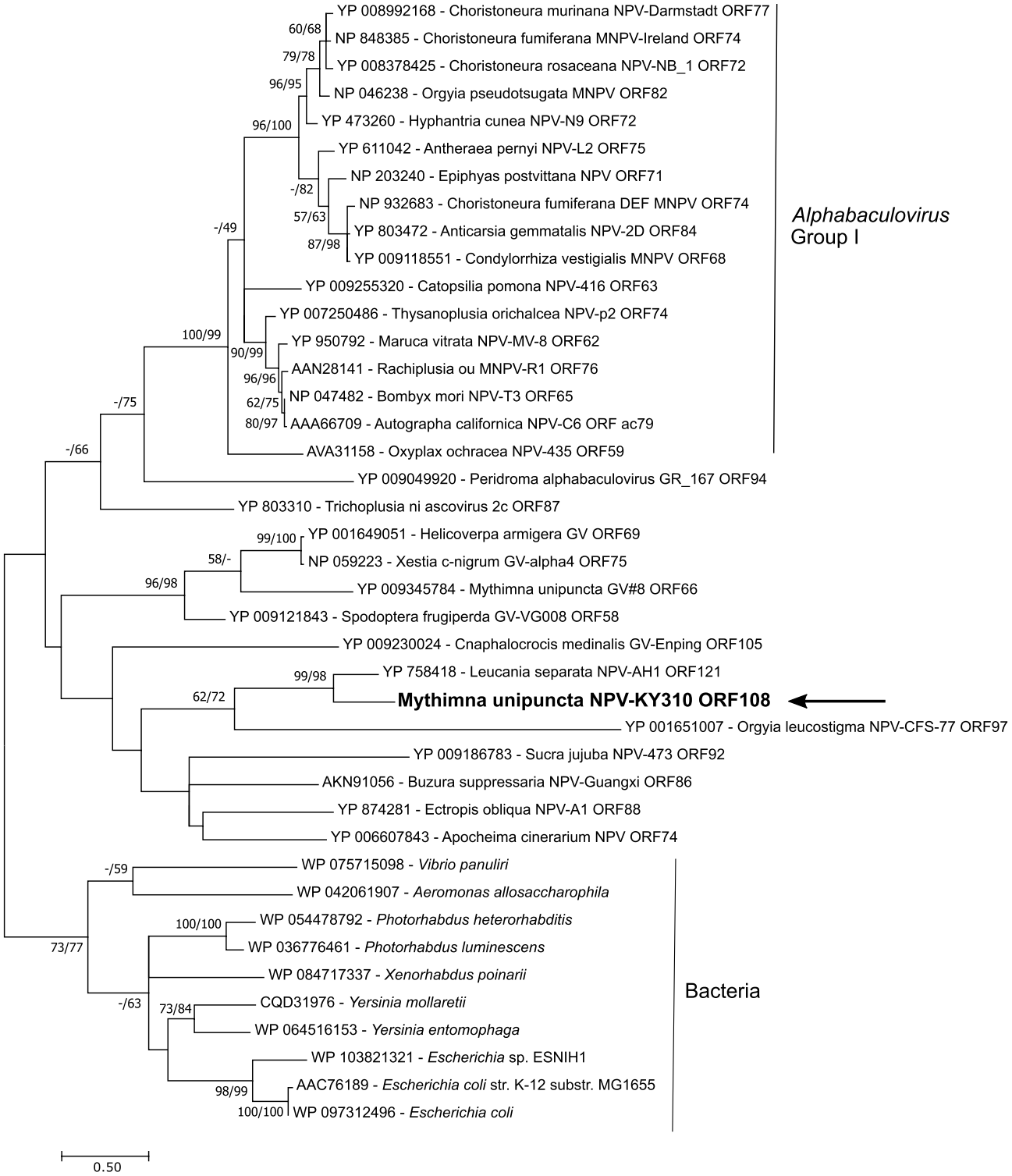

Fig. 7 Phylogeny of baculovirus and bacterial protein sequences containing GIY-YIG nuclease motifs. ML and ME phylograms were inferred from the alignment of amino acid sequences encoded by alphabaculoviruses (NPV), betabaculovirus (GV), ascovirus, and bacterial ORFs, with GenBank accession numbers displayed for each sequence. The ML tree is shown with bootstrap values of ≥ 50% for the ME and ML trees (ME/ML) where they occur, as described for Fig. 4. Lines indicate the positions of group I alphabaculovirus and bacterial sequences, and the MyunNPV-KY310 GIY-YIG nuclease homolog (ORF108) is indicated with an arrow

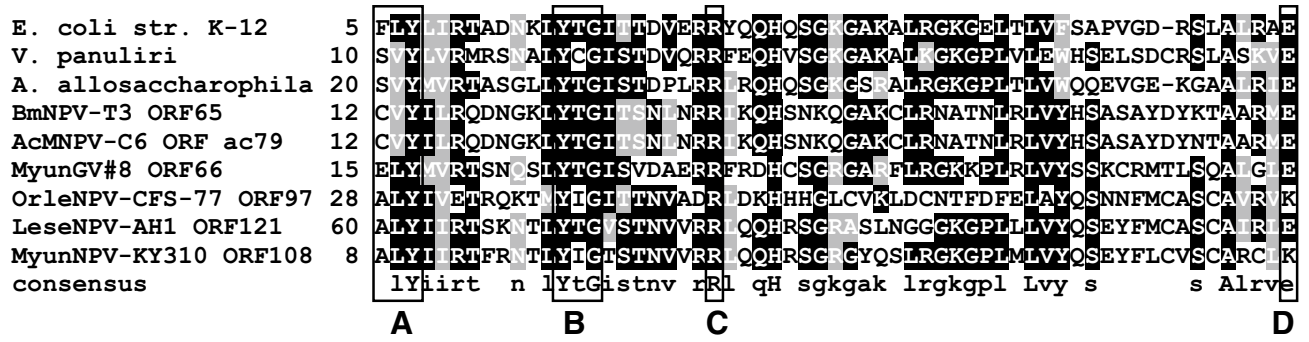

Fig. 8 Alignment of GIY-YIG nuclease domains from selected baculovirus and bacterial sequences. Boxes indicate the locations of GIY (**a**) and YIG (**b**) motifs, as well as conserved arginine (**c**) and glutamate (**d**) residues. Identical amino acids occupying ≥ 50% of aligned positions are shaded in black, and chemically similar amino acids are shaded in gray. Amino acids in the consensus sequence are denoted by uppercase letters for positions in the alignment consisting of the same residue in all the sequences, and lowercase letters for positions in the alignment in which a majority of the sequences consist of the same residue

deletion phenotype characterized by either no impact or a minor impact on viral replication [44, 45]. Although several GIY-YIG nuclease homologs have been identified in group I alphabaculoviruses, the origin of the homolog encoded by ORF108 and its relationship to the group I alphabaculovirus homologs could not be discerned from phylogeny (Fig. 7). GIY-YIG domain-containing proteins have a wide variety of functions involving DNA cleavage, including DNA repair [37]. Inactivation of the AcMNPV GIY-YIG homolog (*ac79*) resulted in reduced titers of infectious budded virus [38]. In contrast, inactivation of the *Bombyx mori* nucleopolyhedrovirus (BmNPV) homolog, Bm65, resulted in the loss of infectious budded virus production [46]. Additional studies have suggested that Bm65 is involved in repair of UV damage of DNA [47, 48]. It is unclear if the ORF108 sequence encodes an active nuclease, as a conserved glutamate residue involved in metal-binding and catalytic activity [49] has been replaced in the ORF108 amino acid sequence with a lysine (Fig. 8).

The genome of MyunNPV-KY310 is the fourth baculovirus genome overall, and the second alphabaculovirus genome, to be determined from a virus of the true armyworm, *M. unipuncta*, along with PsunGV-H, MyunGV#8, and MyunNPV#7. Most baculoviruses possess a narrow host range consisting of one or a few related host species, and there has been no report of the *M. unipuncta* baculoviruses identified from insects other than *M. unipuncta*. However, the possibility that these viruses have a broader host range cannot be excluded. Nevertheless, these virus isolates originate from the same host, and therefore an examination and comparison of their genomes could be expected to provide some insight into the molecular basis for baculovirus host range and specificity. ORFs unique to these viruses, or at least possessing a high degree of sequence similarity, might function to optimize infection and replication in true armyworm larvae or related species such as *M. separata*. Four ORFs of MyunGV#8-22 (*nrk-1*), 41, 56, and 136—grouped in phylogenies with homologous ORFs present in LeseNPV-AH1 [7]. Only two of these ORFS-22 (*nrk-1*) and 56—are also found in MyunNPV-KY310, and the MyunNPV#7 genome only contains a homolog of *nrk-1* [8]. The sequence identity between the MyunNPV-KY310 NRK1 and the homologous sequences of MyunGV#8 and LeseNPV-AH1, as assessed by BLASTp, are 62% and 51.4%, respectively. However, sequence identity between the NRK1 sequences of the two MyunNPV isolates KY310 and #8 is only 28.1%, which suggests that the NRK1 sequences found in the MyunGV#8/LeseNPV-AH1/MyunNPV-KY310 lineage may not specify infectivity towards larvae of genus *Mythimna*. No other ORFs unique to the two MyunNPV genome sequences were identified.

Thézé et al. [34] identified an ORF, *xc138*, from XecnGV with highly conserved homologs in the PsunGV-H and Mythimna separata entomopoxvirus (MySEV) genomes. The authors speculated that the *xc138* gene product might confer specific virulence towards *M. separata* or related hosts. However, no *xc138* homolog was identified in LeseNPV-AH1 [23], MyunGV#8 [7], MyunNPV#7 [8], or MyunNPV-KY310 (Supplementary Table 2). BLASTp and tBLASTn queries with the amino acid sequence encoded by *xc138* failed to find homologous sequences in these genomes.

Other studies have shown that the baculovirus genes *helicase* (*p143*), *late expression factor-7* (*lef-7*), *host range factor-1* (*hrf-1*), *host cell-specific factor-1* (*hcf-1*), immediate-early gene *ie-2*, and *Hyphantria cunea* nucleopolyhedrovirus gene *ep32* also can influence host range and species-specific virulence [50–56]. Five of these genes (*ie-2, hrf-1, hcf-1, lef-7*, and *ep32*) are missing from one or both of the MyunNPV genomes. The *helicase* gene is a baculovirus core gene that is present in all baculovirus genomes. Recombination between the *helicase* genes of the closely

related alphabaculoviruses AcMNPV and *Bombyx mori* nucleopolyhedrovirus (BmNPV) resulted in recombinant isolates of AcMNPV able to infect and fully replicate in *Bombyx mori* larvae and a *B. mori* cell line which were otherwise only semi-permissive for the parental AcMNPV isolate [50]. Further research found that specific substitutions at two positions within the AcMNPV helicase were necessary for AcMNPV to infect and kill *B. mori* larvae [57]. In helicase amino acid alignments, the amino acids at these two positions were not conserved among the helicases of the four true armyworm viruses. In another study, the *helicase* gene of *Trichoplusia ni* granulovirus (TnGV) was unable to support replication of AcMNPV in *Trichoplusia ni* cells when it was used to replace the native AcMNPV *helicase* sequence in a bacmid, even though both parental viruses can infect and replicate in *T. ni* [58]. The helicases of AcMNPV and TnGV share only 29% sequence identity, and the inability of the TnGV helicase to support replication of AcMNPV may represent the failure of TnGV helicase to assemble with other AcMNPV proteins in a presumptive baculovirus replication complex. Similarly, the helicases of the two MyunNPV isolates exhibit 41.2% sequence identity, while the helicases of AcMNPV and BmNPV share 96% sequence identity. Together, these observations collectively suggest that there is not a single conserved helicase amino acid sequence required for infectivity against true armyworm larvae.

The ability to avoid or overcome host apoptosis is key to the establishment of baculovirus infection, and has been shown to shape baculovirus host range [59–62]. Both the KY310 and #7 isolates of MyunNPV encode two inhibitor-of-apoptosis proteins, one each from the IAP-2 and IAP-3 lineages [63]. However, the IAP-2 sequences of the two isolates only share 25.4% sequence identity, and the MyunNPV#7 IAP-3 does not appear among the matches of a BLASTp query with the MyunNPV-KY310 IAP-3 sequence. The MyunNPV-KY310 genome contains a *p49* ORF, which encodes a homolog of the P35 anti-apoptosis protein [23, 64]. In contrast, the MyunNPV#7 genome contains no *p35* homolog. The MyunGV#8 and PsunGV-H genomes each contain a single copy of a gene coding IAP-5 and no *p35* homolog. These observations suggest that successful evasion of the *M. unipuncta* larval defensive apoptotic response does not require that a baculovirus produce a specific array of anti-apoptotic proteins, or that it produces anti-apoptotic proteins that display much sequence identity with each other.

The genome sequence determination of a fourth baculovirus from *M. unipuncta* and its comparison to the other *M. unipuncta* baculovirus genome sequences has confirmed the remarkable extent to which baculoviruses can diverge, yet still infect and replicate in the same host species. This study also alludes to the complex nature of baculovirus host range and specificity. Much work remains to elucidate the details of its genetic and molecular basis.

Acknowledgements The authors would like to thank Pallavi Thapa, Daniel Barakh, and Anita Ghosh (USDA-ARS, Beltsille, MD) for assistance with dideoxy sequencing, and Doug Baldwin (AAFC) for assistance with assembly of 454 sequence reads and initial annotation of the genome. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Author contributions Conceived and designed the experiments: RLH, JDM, GRB, DAT, MAE. Performed the experiments: RLH, JDM, GRB, MAE. Analyzed the data: RLH. Wrote the paper: RLH, JDM, GRB, DAT, MAE.

Compliance with ethical standards

Conflict of interest The authors have no conflicts of interest to declare.

Ethical approval The research described in this paper does not use any human or animal subjects.

References

1. Herniou EA, Arif BM, Becnel JJ, Blissard GW, Bonning B, Harrison RL, Jehle JA, Theilmann DA, Vlak JM (2011) Baculoviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (eds) Virus taxonomy: ninth report of the international committee on taxonomy of viruses. Elsevier, Oxford, pp 163–174
2. Harrison RL, Hoover K (2012) Baculoviruses and other occluded insect viruses. In: Vega FE, Kaya HK (eds) Insect pathology, 2nd edn. Academic Press, Boston, pp 73–131
3. Rohrmann GF (2013) Baculovirus molecular biology, 3rd edn. National Center for Biotechnology Information (US), Bethesda
4. van Oers MM, Pijlman GP, Vlak JM (2015) Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. J Gen Virol 96:6–23. https://doi.org/10.1099/vir.0.067108-0
5. Lacey LA, Grzywacz D, Shapiro-Ilan DI, Frutos R, Brownbridge M, Goettel MS (2015) Insect pathogens as biological control agents: back to the future. J Invertebr Pathol 132:1–41. https://doi.org/10.1016/j.jip.2015.07.009 pii]
6. Keathley CP, Harrison RL, Potter DA (2012) Baculovirus infection of the armyworm (Lepidoptera: Noctuidae) feeding on spiny- or smooth-edged grass (*Festuca* spp.) leaf blades. Biol Control 61:147–154
7. Harrison RL, Rowley DL, Mowery J, Bauchan GR, Theilmann DA, Rohrmann GF, Erlandson MA (2017) The complete genome sequence of a second distinct betabaculovirus from the true armyworm, *Mythimna unipuncta*. PLoS ONE 12(1):e0170510. https://doi.org/10.1371/journal.pone.0170510
8. Harrison RL, Mowery JD, Rowley DL, Bauchan GR, Theilmann DA, Rohrmann GF, Erlandson MA (2018) The complete genome sequence of a third distinct baculovirus isolated from the true armyworm, *Mythimna unipuncta*, contains two copies of the lef-7 gene. Virus Genes 54:297–310. https://doi.org/10.1007/s11262-017-1525-0
9. Harrison RL, Puttler B, Popham HJ (2008) Genomic sequence analysis of a fast-killing isolate of *Spodoptera frugiperda*

multiple nucleopolyhedrovirus. J Gen Virol 89:775–790. https://doi.org/10.1099/vir.0.83566-0
10. Harrison RL, Keena MA, Rowley DL (2014) Classification, genetic variation and pathogenicity of *Lymantria dispar* nucleopolyhedrovirus isolates from Asia, Europe, and North America. J Invertebr Pathol 116:27–35. https://doi.org/10.1016/j.jip.2013.12.005
11. Guo F-B, Zhang C-T (2006) ZCURVE_V: a new self-training system for recognizing protein-coding genes in viral and phage genomes. BMC Bioinform 7:9. https://doi.org/10.1186/1471-2105-7-9
12. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V (2018) A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 430(15):2237–2243. https://doi.org/10.1016/j.jmb.2017.12.007
13. Hu ZH, Arif BM, Jin F, Martens JW, Chen XW, Sun JS, Zuidema D, Goldbach RW, Vlak JM (1998) Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. J Gen Virol 79:2841–2851
14. Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE 5(6):e11147. https://doi.org/10.1371/journal.pone.0011147
15. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 5:113. https://doi.org/10.1186/1471-2105-5-113
16. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series, vol 41, pp 95–98
17. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
18. Silvestro D, Michalak I (2012) raxmlGUI: a graphical front-end for RAxML. Org Divers Evol 12:335–337
19. Jehle JA, Lange M, Wang H, Hu Z, Wang Y, Hauschild R (2006) Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. Virology 346:180–193. https://doi.org/10.1016/j.virol.2005.10.032
20. Ackermann H-W, Smirnoff WA (1983) A morphological investigation of 23 baculoviruses. J Invertebr Pathol 41:269–280
21. Garavaglia MJ, Miele SA, Iserte JA, Belaich MN, Ghiringhelli PD (2012) The ac53, ac78, ac101, and ac103 genes are newly discovered core genes in the family Baculoviridae. J Virol 86(22):12069–12079. https://doi.org/10.1128/JVI.01873-12
22. Javed MA, Biswas S, Willis LG, Harris S, Pritchard C, van Oers MM, Donly BC, Erlandson MA, Hegedus DD, Theilmann DA (2017) *Autographa californica* multiple nucleopolyhedrovirus AC83 is a per os infectivity factor (PIF) protein required for occlusion-derived virus (ODV) and budded virus nucleocapsid assembly as well as assembly of the PIF complex in ODV envelopes. J Virol. https://doi.org/10.1128/JVI.02115-16
23. Xiao H, Qi Y (2007) Genome sequence of *Leucania seperata* nucleopolyhedrovirus. Virus Genes 35:845–856
24. Zanotto PM, Kessing BD, Maruniak JE (1993) Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. J Invertebr Pathol 62:147–164
25. Herniou EA, Jehle JA (2007) Baculovirus phylogeny and evolution. Curr Drug Targets 8:1043–1050
26. Breitenbach JE, El-Sheikh el SA, Harrison RL, Rowley DL, Sparks ME, Gundersen-Rindal DE, Popham HJ (2013) Determination and analysis of the genome sequence of *Spodoptera littoralis* multiple nucleopolyhedrovirus. Virus Res 171:194–208. https://doi.org/10.1016/j.virusres.2012.11.016
27. Pang Y, Yu J, Wang L, Hu X, Bao W, Li G, Chen C, Han H, Hu S, Yang H (2001) Sequence analysis of the *Spodoptera litura* multicapsid nucleopolyhedrovirus genome. Virology 287:391–404
28. Wang JW, Qi YP, Huang YX, Li SD (1995) Nucleotide sequence of a 1446 base pair *Sal*I fragment and structure of a novel early gene of *Leucania separata* nuclear polyhedrosis virus. Arch Virol 140:2283–2291
29. Jin T, Qi Y, Qi B, Jin H, Huang Y (1997) Nucleotide sequence of a 5423 base pairs fragment of the LsNPV genome and comparison with the AcNPV genome. Virus Genes 14:201–210
30. Thézé J, Lopez-Vaamonde C, Cory JS, Herniou EA (2018) Biodiversity, evolution, and ecological specialization of baculoviruses: a treasure trove for future applied research. Viruses 10:366
31. Bideshi DK, Renault S, Stasiak K, Federici BA, Bigot Y (2003) 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 84:2531–2544
32. Volkoff AN, Beliveau C, Rocher J, Hilgarth R, Levasseur A, Duonor-Cerutti M, Cusson M, Webb BA (2002) Evidence for a conserved polydnavirus gene family: ichnovirus homologs of the CsIV repeat element genes. Virology 300:316–331
33. Tanaka K, Lapointe R, Barney WE, Makkay AM, Stoltz D, Cusson M, Webb BA (2007) Shared and species-specific features among ichnovirus genomes. Virology 363:26–35
34. Theze J, Takatsuka J, Nakai M, Arif B, Herniou EA (2015) Gene acquisition convergence between entomopoxviruses and baculoviruses. Viruses 7:1960–1974. https://doi.org/10.3390/v7041960
35. Jakubowska AK, Peters SA, Ziemnicka J, Vlak JM, van Oers MM (2006) Genome sequence of an enhancin gene-rich nucleopolyhedrovirus (NPV) from *Agrotis segetum*: collinearity with *Spodoptera exigua* multiple NPV. J Gen Virol 87:537–551
36. Becker D, Knebel-Morsdorf D (1993) Sequence and temporal appearance of the early transcribed baculovirus gene HE65. J Virol 67:5867–5872
37. Dunin-Horkawicz S, Feder M, Bujnicki JM (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. BMC Genom 7:98. https://doi.org/10.1186/1471-2164-7-98
38. Wu W, Passarelli AL (2012) The *Autographa californica* M nucleopolyhedrovirus ac79 gene encodes an early gene product with structural similarities to UvrC and intron-encoded endonucleases that is required for efficient budded virus production. J Virol 86:5614–5625. https://doi.org/10.1128/JVI.06252-11
39. Ardisson-Araujo DM, Lima RN, Melo FL, Clem RJ, Huang N, Bao SN, Sosa-Gomez DR, Ribeiro BM (2016) Genome sequence of *Perigonia lusca* single nucleopolyhedrovirus: insights into the evolution of a nucleotide metabolism enzyme in the family Baculoviridae. Sci Rep 6:24612. https://doi.org/10.1038/srep24612
40. Wang J, Zhu Z, Zhang L, Hou D, Wang M, Arif B, Kou Z, Wang H, Deng F, Hu Z (2016) Genome sequencing and analysis of *Catopsilia pomona* nucleopolyhedrovirus: a distinct species in group I *Alphabaculovirus*. PLoS ONE 11(5):e0155134. https://doi.org/10.1371/journal.pone.0155134
41. Harrison RL, Rowley DL, Mowery JD, Bauchan GR, Burand JP (2017) The *Operophtera brumata* nucleopolyhedrovirus (OpbuNPV) represents an early, divergent lineage within genus *Alphabaculovirus*. Viruses. https://doi.org/10.3390/v9100307
42. Santos ER, Oliveira LB, Peterson L, Sosa-Gomez DR, Ribeiro BM, Ardisson-Araujo DMP (2018) The complete genome sequence of the first hesperiid-infecting alphabaculovirus isolated from the leguminous pest *Urbanus proteus* (Lepidoptera: Hesperiidae). Virus Res 249:76–84. https://doi.org/10.1016/j.virusres.2018.03.009
43. Ohkawa T, Rowe AR, Volkman LE (2002) Identification of six *Autographa californica* multicapsid nucleopolyhedrovirus early genes that mediate nuclear localization of G-actin. J Virol 76:12281–12289
44. Gandhi KM, Ohkawa T, Welch MD, Volkman LE (2012) Nuclear localization of actin requires AC102 in *Autographa californica*

multiple nucleopolyhedrovirus-infected cells. J Gen Virol 93:1795–1803. https://doi.org/10.1099/vir.0.041848-0
45. Ono C, Nakatsukasa T, Nishijima Y, Asano S, Sahara K, Bando H (2007) Construction of the BmNPV T3 bacmid system and its application to the functional analysis of BmNPV *he65*. J Insect Biotechnol Sericol 76:161–167
46. Tang Q, Li G, Yao Q, Chen L, Feng F, Yuan Y, Chen K (2013) Bm65 is essential for the propagation of *Bombyx mori* nucleopolyhedrovirus. Curr Microbiol 66:22–29. https://doi.org/10.1007/s00284-012-0236-y
47. Tang Q, Hu Z, Yang Y, Wu H, Qiu L, Chen K, Li G (2015) Overexpression of Bm65 correlates with reduced susceptibility to inactivation by UV light. J Invertebr Pathol 127:87–92. https://doi.org/10.1016/j.jip.2015.03.003
48. Tang Q, Wu P, Hu Z, Yang Y, Qiu L, Liu H, Zhu S, Guo Z, Xia H, Chen K, Li G (2017) Evidence for the role of BmNPV Bm65 protein in the repair of ultraviolet-induced DNA damage. J Invertebr Pathol 149:82–86. https://doi.org/10.1016/j.jip.2017.08.004
49. Mak AN, Lambert AR, Stoddard BL (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. Structure 18:1321–1331. https://doi.org/10.1016/j.str.2010.07.006
50. Croizier G, Croizier L, Argaud O, Poudevigne D (1994) Extension of *Autographa californica* nuclear polyhedrosis virus host range by interspecific replacement of a short DNA sequence in the p143 helicase gene. Proc Natl Acad Sci USA 91:48–52
51. Lu A, Miller LK (1995) Differential requirements for baculovirus late expression factor genes in two cell lines. J Virol 69:6265–6272
52. Chen CJ, Thiem SM (1997) Differential infectivity of two *Autographa californica* nucleopolyhedrovirus mutants on three permissive cell lines is the result of lef-7 deletion. Virology 227:88–95
53. Chen CJ, Quentin ME, Brennan LA, Kukel C, Thiem SM (1998) *Lymantria dispar* nucleopolyhedrovirus hrf-1 expands the larval host range of *Autographa californica* nucleopolyhedrovirus. J Virol 72:2526–2531
54. Lu A, Miller LK (1996) Species-specific effects of the hcf-1 gene on baculovirus virulence. J Virol 70:5123–5130
55. Prikhod'ko EA, Lu A, Wilson JA, Miller LK (1999) In vivo and in vitro analysis of baculovirus ie-2 mutants. J Virol 73:2460–2468
56. Shirata N, Ikeda M, Kobayashi M (2010) Identification of a *Hyphantria cunea* nucleopolyhedrovirus (NPV) gene that is involved in global protein synthesis shutdown and restricted *Bombyx mori* NPV multiplication in a *B. mori* cell line. Virology 398:149–157. https://doi.org/10.1016/j.virol.2009.11.049
57. Argaud O, Croizier L, Lopez-Ferber M, Croizier G (1998) Two key mutations in the host-range specificity domain of the p143 gene of *Autographa californica* nucleopolyhedrovirus are required to kill *Bombyx mori* larvae. J Gen Virol 79:931–935
58. Bideshi DK, Federici BA (2000) The *Trichoplusia ni* granulovirus helicase is unable to support replication of *Autographa californica* multicapsid nucleopolyhedrovirus in cells and larvae of *T. ni*. J Gen Virol 81:1593–1599
59. Clem RJ, Fechheimer M, Miller LK (1991) Prevention of apoptosis by a baculovirus gene during infection of insect cells. Science 254:1388–1390
60. Griffiths CM, Barnett AL, Ayres MD, Windass J, King LA, Possee RD (1999) In vitro host range of *Autographa californica* nucleopolyhedrovirus recombinants lacking functional p35, iap1 or iap2. J Gen Virol 80:1055–1066
61. Zhang P, Yang K, Dai X, Pang Y, Su D (2002) Infection of wild-type *Autographa californica* multicapsid nucleopolyhedrovirus induces in vivo apoptosis of *Spodoptera litura* larvae. J Gen Virol 83:3003–3011
62. Liu B, Becnel JJ, Zhang Y, Zhou L (2011) Induction of reaper ortholog mx in mosquito midgut cells following baculovirus infection. Cell Death Differ 18:1337–1345. https://doi.org/10.1038/cdd.2011.8
63. Clem RJ (2015) Viral IAPs, then and now. Semin Cell Dev Biol 39:72–79. https://doi.org/10.1016/j.semcdb.2015.01.011 pii]
64. Du Q, Lehavi D, Faktor O, Qi Y, Chejanovsky N (1999) Isolation of an apoptosis suppressor gene of the *Spodoptera littoralis* nucleopolyhedrovirus. J Virol 73:1278–1285