Structural divergence among genomes of closely related baculoviruses and its implications for baculovirus evolution

Robert L. Harrison *

Invasive Insect Biocontrol and Behavior Laboratory, USDA Agricultural Research Service, PSI, Building 011A, Room 214, BARC-W, 10300 Baltimore Ave., Beltsville, MD 20705, USA

A R T I C L E   I N F O

Article history:
Received 12 February 2009
Accepted 5 March 2009
Available online 19 May 2009

Keywords:
Baculovirus
Nucleopolyhedrovirus
Granulovirus
Autographa californica MNPV
Rachiplusia ou MNPV
Plutella xylostella MNPV
Xestia c-nigrum GV
Helicoverpa armigera GV
Recombination
Genomes

A B S T R A C T

Baculoviruses are members of a large, well-characterized family of dsDNA viruses that have been identified from insects of the orders Lepidoptera, Hymenoptera, and Diptera. Baculovirus genomes from different virus species generally exhibit a considerable degree of structural diversity. However, some sequenced baculovirus genomes from closely related viruses are structurally very similar and share overall nucleotide sequence identities in excess of 95%. This review focuses on the comparative analysis of partial and complete nucleotide sequences from two groups of closely related baculoviruses with broad host ranges: (a) group I multiple nucleopolyhedroviruses (MNPVs) from a cluster including Autographa californica (Ac)MNPV, Rachiplusia ou (Ro)MNPV, and Plutella xylostella (Pxy)MNPV; and (b) granuloviruses (GVs) from a cluster including Xestia c-nigrum (Xecn)GV and Helicoverpa armigera (Hear)GV. Even though the individual viruses in these clusters share high nucleotide sequence identities, a significant degree of genomic rearrangement (in the form of insertions, deletions, and homologous recombination resulting in allelic replacement) is evident from alignments of their genomes. These observations suggest an important role for recombination in the early evolution and biological characteristics of baculoviruses of these two groups.

Published by Elsevier Inc.

1. Introduction: 50 complete baculovirus genome sequences and counting...

Invertebrate viruses of family Baculoviridae have been studied extensively and developed as biopesticides and as vectors for gene delivery and recombinant protein production in insect and mammalian cells (Kost et al., 2005; Moscardi, 1999; Summers, 2006; van Beek and Davis, 2007). The genomes of these viruses are large (80–180 kb), but since the reported sequence of the first baculovirus genome in 1994, more and more baculovirus genomes are being completely sequenced and analyzed. As of December 2008, there were 50 complete baculovirus genome sequences on file at Genbank http://www.ncbi.nlm.nih.gov/genomes/genlist.cgi?taxid=10239&type=5&name=Viruses. Analyses of many of these genome sequences have been published in the scientific literature (van Oers and Vlak, 2007).

All baculovirus genomes consist of a single, circular double-stranded DNA molecule that varies widely in size and nucleotide distribution from baculovirus to baculovirus (van Oers and Vlak, 2007). This molecule is densely packed with open reading frames (ORFs) on both strands and in both orientations. All published baculovirus genome sequences carry 30 core genes (McCarthy and Theilmann, 2008; van Oers and Vlak, 2007) along with many other ORFs found either on a subset of genomes or unique to a particular genome. Differences in the sequence, distribution, order, and orientation of the core genes and the hundreds of additional ORFs identified on baculovirus genomes delineate individual baculovirus species and groups of related baculoviruses.

With the availability of whole baculovirus genomes, it has been possible to construct a classification and evolutionary history of these viruses in a robust fashion through phylogenomics (Herniou and Jehle, 2007; Herniou et al., 2001; Herniou et al., 2003). The lepidopteran nucleopolyhedroviruses (NPs), lepidopteran granuloviruses (GVs), and dipteran and hymenopteran baculoviruses form four distinct clades, and a new classification of baculoviruses in which these four clades are elevated to genera has been proposed (Jehle et al., 2006a). Most of the baculovirus genomes that have been determined derive from lepidopteran NPVs. A finer degree of organization among these viruses can be discerned, with lepidopteran NPVs divided into two groups (group I and group II), and two distinct sub-groups evident among group I viruses (clade 1a and clade 1b) (Jehle et al., 2006b; Zanotto et al., 1993).

The baculovirus genomes that have been determined to date mostly consist of sequences from viruses that have diverged extensively from each other. However, some genome sequences have been reported from closely related baculoviruses with nucleotide sequence identities over 95%. Table 1 shows pairwise comparisons...
of closely related group I and group II NPVs with nucleotide sequence identities ranging from 97.0% to 99.9%. Multiple substitutions and small insertions and deletions (indels) distinguish these pairs of viruses, with some larger indels observed for some of these NPV pairs. In general, the substitutions and indels are concentrated in or around homologous repeat regions (hrs) and baculovirus repeated ORF (bro) genes. The hrs are intergenic sequences consisting of repeated sequences. In tissue culture, these hrs can act as cis-acting enhancers of baculovirus gene transcription (Guarino et al., 1986; Guarino and Summers, 1986) and as origins of viral DNA replication (Hilton and Winstanley, 2007; Pearson et al., 1992). The bro genes constitute a widespread multigene family found in baculoviruses and other invertebrate dsDNA viruses (Bideshi et al., 2003). While a number of activities and putative functions have been characterized for bro gene products, their role in the baculovirus life cycle is not known with precision (Kang et al., 2004; 1999; 2003; Zemskov et al., 2000). Both hrs and bro gene sequences have been found near regions of variability in ORF content among baculovirus genomes, suggesting that these sequences are involved in baculovirus gene loss and acquisition (de Jong et al., 2005; Hayakawa et al., 2000; Li et al., 2002).

This review focuses on the comparative genomics of two different groups of closely related baculoviruses: (a) multiple nucleopolyhedroviruses (MNPVs) from a cluster including Autographa californica (Ac)MNPV, Rachiplusia ou (Ro)MNPV, and Plutella xylostella (Pxy)MNPV; and (b) granuloviruses (GVs) from a cluster including Xestia c-nigrum (Xc)GV and Helicoverpa armigera (Hear)GV. These groups of viruses are notable in having remarkably broad host ranges relative to other baculoviruses. Even though the genomes of viruses in these groups exhibit high nucleotide sequence identities, a relatively high degree of genomic rearrangement is evident from alignments of completely sequenced genomes from these groups. The extent of rearrangement that has occurred in the different lineages of these viruses appears to be greater than that of the genomes listed in Table 1, suggesting a significant role for recombination in the evolution of these viruses.

2. “AcMNPV-like” viruses: AcMNPV, RoMNPV, PxyMNPV-CL3

The term “AcMNPV-like” describes viruses that are similar to AcMNPV, the types species for Baculoviridae (Theilmann et al., 2005), a model for studies on baculovirus molecular biology, and a frequently-used vector for recombinant protein production in the baculovirus-insect cell expression system. These viruses occur in clade 1a of the group I lepidopteran NPVs (Jehle et al., 2006b). AcMNPV has been reported to infect up to 43 lepidopteran species from 11 families (Payne, 1986), and variants of this virus have been isolated from many different host species (Brown et al., 1985; Jehle et al., 2006b; Maeda et al., 1990; Miller and Dawes, 1978; Smith and Summers, 1979; Yanase et al., 2000). The C6 clone of AcMNPV was the first baculovirus to have its genome sequenced (Ayres et al., 1994).

Complete genome sequences have been determined for two other AcMNPV-like viruses. R. ou MNPV was isolated from a population of infected mint loopers in 1960 (Paschke and Hamm, 1961; Paschke and Sweet, 1966). A variant of this virus was isolated from a different host, Anagropa fulciüera, in 1985 (Harrison and Bonning, 1999; Hostetter and Puttler, 1991). These viruses also can infect many species of Lepidoptera, including 31 species from 10 families, and are more virulent than AcMNPV against a selection of host species, including Helicoverpa zea and Ostrinia nubilalis (Harrison and Bonning, 1999; Hostetter and Puttler, 1991). The sequence of the RoMNPV isolate R1 was published in 2003 (Harrison and Bonning, 2003). P. xylostella MNPV, clone CL3, was plaque-purified from a mixed sample of GV and NPV isolated from cadavers of P. xylostella (diamondback moth) larvae (Kariuki and Mcintosh, 1999). This virus is notable for being significantly more virulent against diamondback moth larvae than AcMNPV (Farrar et al., 2004; Kariuki and Mcintosh, 1999). Though PxyMNPV-CL3 has not been tested against a wide range of species, comparisons of PxyMNPV-CL3 and AcMNPV restriction endonuclease fragment profiles indicated that PxyMNPV-CL3 likely is an AcMNPV variant (Kariuki et al., 2000). This relationship was confirmed when its genome sequence was determined and published in 2007 (Harrison and Lynn, 2007).

The genome sequences of AcMNPV-C6, RoMNPV-R1, and PxyMNPV-CL3 exhibit minor variations in genome size, nucleotide distribution, and number of ORFs present (Harrison and Bonning, 2003; Harrison and Lynn, 2007). All three viruses have 9 hrs in the same genomic positions, with differences in the number of repeats at each region observed among the viruses. PxyMNPV-CL3 displays high nucleotide sequence and ORF amino acid sequence identities with AcMNPV-C6, with percentages of 98.6 and 98.7 for those categories. RoMNPV-R1 has somewhat lower sequence identities of 95.7% and 95.4%, respectively. Phylogenetic inference and Kimura-2-parameter nucleotide distances for the core genes lef-8, lef-9, and polh clearly indicate that PxyMNPV-CL3 is a variant of AcMNPV. An argument has been made that AcMNPV/RoMNPV-R1 and AcMNPV should be considered as separate species on the basis of host range differences and Kimura-2-parameter nucleotide distances with AcMNPV isolates that often exceed 0.050 substitutions/site (Jehle et al., 2006b).

Although the genomes of AcMNPV-C6, RoMNPV-R1, and PxyMNPV-CL3 are extensively co-linear, there are a few differences in ORF distribution among the three genomes. AcMNPV-C6 has a single bro gene, which also can be found in PxyMNPV-CL3. This gene, along with an adjacent ORF encoding a conotoxin-like gene product (ctl), was apparently lost by the RoMNPV/AFMNPV lineage as it diverged from AcMNPV (Fig. 1A; Federici and Hice, 1997; Harrison and Bonning, 1999). PxyMNPV-CL3 and an AcMNPV variant isolated from Spodoptera litera (SpinMNPV-OM1) have acquired an additional bro gene not found in AcMNPV-C6 or RoMNPV (Fig. 1B). Changes in the bro gene distribution have been frequently observed among baculovirus genomes that are more extensively diverged than the group I NPVs. While it appears that at least one bro gene is essential for replication of Bombyx mori NPV, the AcMNPV ORF2 bro gene is not necessary for replication (Bideshi et al., 2003). It is not clear how the different number and arrangements of bro genes among AcMNPV, RoMNPV, and PxyMNPV-CL3 affect baculovirus infection and replication.

PxyMNPV-CL3 was also found to be missing a large portion of ORF ac86 (pnk/pml; Fig. 1C). The ac86 ORF encodes a protein with sequence similarity to, and the enzymatic activities of, two sepa-

Table 1
Pairwise comparisons of highly similar baculovirus genomes.

<table>
<thead>
<tr>
<th>Virus isolates compared</th>
<th>% nt sequence identity</th>
<th>Major structural differences</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Helicoverpa za SNPV vs. H. armigera SNPV (G1)</td>
<td>97.0</td>
<td>Novel 334 bp insert with ORF in HzSNPV</td>
<td>Chen et al. (2002)</td>
</tr>
<tr>
<td>H. armigera SNPV-C4 vs. -G1</td>
<td>98.1</td>
<td>–</td>
<td>Zhang et al. (2005)</td>
</tr>
<tr>
<td>Mamestra configurata NPV-A 90/2 vs. -90/4</td>
<td>99.5</td>
<td>Additional baculovirus repeated of (bro) gene in 90/2</td>
<td>Li et al. (2005)</td>
</tr>
<tr>
<td>Antheraea pernyi MNPV-L vs. -Z</td>
<td>99.9</td>
<td>113 bp indel in egt gene</td>
<td>Fan et al. (2007)</td>
</tr>
<tr>
<td>Spodoptera frugiperda MNPV-3AF2 vs. -19</td>
<td>99.6</td>
<td>1427 bp deletion in egt gene of 3AF2</td>
<td>Wolff et al. (2008)</td>
</tr>
</tbody>
</table>

This review focuses on the comparative genomics of two different groups of closely related baculoviruses: (a) multiple nucleopolyhedroviruses (MNPVs) from a cluster including Autographa californica (Ac)MNPV, Rachiplusia ou (Ro)MNPV, and Plutella xylostella (Pxy)MNPV; and (b) granuloviruses (GVs) from a cluster including Xestia c-nigrum (Xc)GV and Helicoverpa armigera (Hear)GV. These groups of viruses are notable in having remarkably broad host ranges relative to other baculoviruses. Even though the genomes of viruses in these groups exhibit high nucleotide sequence identities, a relatively high degree of genomic rearrangement is evident from alignments of completely sequenced genomes from these groups. The extent of rearrangement that has occurred in the different lineages of these viruses appears to be greater than that of the genomes listed in Table 1, suggesting a significant role for recombination in the evolution of these viruses.
rate phage T4 enzymes: an ATP-dependent RNA ligase and a polynucleotide 5’-kinase/3’-phosphatase (Martins and Shuman, 2004). Eliminating expression of ac86 had no impact on AcMNPV replication in cell culture (Durantel et al., 1998). The ac86 ORF deletion in PlxyMNPV-CL3 occurred in a region adjacent to hr3 that has a variable ORF arrangement among group I NPVs. An AcMNPV plaque isolate, AcMNPV-1.2, also contains a deletion removing most of the ac86 ORF (Durantel et al., 1998). A 1.6 kb sequence from the ac86 region of this plaque isolate was determined. Comparison of this sequence with the corresponding sequence of PlxyMNPV-CL3 revealed that the deletion endpoints are conserved in these two viruses. This region is more strongly conserved between AcMNPV-1.2 and PlxyMNPV-CL3 (98% nt sequence identity) than between AcMNPV-C6 and PlxyMNPV-CL3 (93.1% identity). In contrast to PlxyMNPV-CL3, RoMNPV-R1 has an intact ac86 ORF and shares 99.8% nucleotide sequence identity with AcMNPV-C6 in this region. A neighbor-joining phylogram with these nucleotide sequences illustrates the differences in sequence conservation in this region (Fig. 2A). Kimura-2-parameter nucleotide distances between AcMNPV-C6 and RoMNPV and between AcMNPV-1.2 and PlxyMNPV-CL3 range between 0.002 and 0.015 substitutions/site, but comparing AcMNPV-C6 or RoMNPV with either AcMNPV-1.2 or PlxyMNPV-CL3 yields nucleotide distances ranging from 0.068 to 0.074 substitutions/site. These results suggest that homologous recombination in this region occurred with another group I NPV, resulting in allelic replacement in a lineage of AcMNPV variants that includes PlxyMNPV-CL3 and AcMNPV-1.2.

Homologous recombination and allelic replacement in baculoviruses has been reported previously for the polyhedrin gene of AcMNPV (Jehle, 2004). Using a Hidden Markov Model analysis method for detecting sequences that are mosaics of two different sources, Jehle (2004) showed that the AcMNPV polyhedrin nucleotide sequence appears to be a mosaic of group I sequences recombined with a sequence from a group II NPV (Fig. 2B), possibly an ancestor of the NPVs in the clade containing Trichoplusia ni SNPV, Plusia orichalcea SNPV, and Chrysodeixis chalcites NPV (Herniou and Jehle, 2007). While PlxyMNPV-CL3 and other AcMNPV isolates carry this hybrid polh gene, RoMNPV, AfMNPV and other group I NPVs do not.

The PlxyMNPV-CL3 genome carries a variant form of the ie-2 gene, which is expressed early during infection and encodes a protein involved in both transcriptional regulation and DNA replication (Carson et al., 1988; Carson et al., 1991; Kool et al., 1994; Passarelli and Miller, 1993; Yoo and Guarino, 1994). The closest matches to PlxyMNPV-CL3 IE-2 amino acid sequence in a BLAST search were the AcMNPV and RoMNPV IE-2 sequences (Harrison and Lynn, 2007). Phylogenetic inference of IE-2 amino acid sequence relationships placed the PlxyMNPV-CL3 sequence with other IE-2 sequences of group I/clade 1a viruses. However, the PlxyMNPV-CL3 IE-2 amino acid sequence only shares 37% identity with the AcMNPV IE-2 sequence, and the branch lengths indicate that PlxyMNPV-CL3 IE-2 is only distantly related to the IE-2 sequences of other clade 1a NPVs. An alignment of PlxyMNPV-CL3 and AcMNPV-C6 nucleotide sequences in the region of the ie-2 gene clearly indicate that PlxyMNPV-CL3 obtained its copy of the ie-2 gene by recombination with another currently undescribed NPV, resulting in allelic replacement of the ie-2 sequence. This can be seen by the sharp decrease in sequence identity in the region containing the ie-2 coding sequence, from over 93% with no or minimal gaps required for an optimal alignment, to 67.4% with many gaps required (Fig. 2C). The AcMNPV ie-2 gene is required for optimal levels of late promoter activation and hr-mediated plasmid

---

**Fig. 1.** ORF maps showing the differences among group I/clade 1a NPVs in (A) the distribution of bro ORFs and (B) the occurrence of ORF ac86 (pnk/pnl).
3. X. c-nigrum and H. armigera granuloviruses

The GV of the spotted cutworm, X. c-nigrum, has the largest genome determined so far for a baculovirus at 178,733 nt. X. c-nigrum (Xecn)GV is part of a group of closely related GVs isolated from noctuid moths, including Autographa gamma GV, Hoplodrina ambigua GV, Euxoa ochrogaster GV, and Scotogramma trifoli GV (Lange et al., 2004). Kimura-2-parameter nucleotide distances for lef-8, lef-9, and granulin among these granuloviruses suggest that they are iso-

lates of the same species (Jehle et al., 2006a). In addition to these viruses, four GVs reported from four additional noctuid moths (Hydraecia amurensis, Caleana leucostigma, Aletia pallens and Pseudaletia separata) were found to have restriction endonuclease di-
gest profiles that matched or were very similar to that of XecnGV (Goto et al., 1992). Taken together, these GV isolates represent a baculovirus species with an exceptionally broad host range encompassing at least 10 different hosts.

A granulovirus from the Old World bollworm, H. armigera, was isolated in 1971 and found to be capable of infecting and killing larvae of three species of noctuid moths in addition to H. armigera (Hamm, 1982; Whitlock, 1974). A 3.2 kbp fragment of HearGV containing an enhancer gene was sequenced (Roelvink et al., 1995); this fragment’s nucleotide sequence was 84.8% identical (excluding gaps) with a region in the XecnGV genome containing the enhan-
cin-3 gene. When the entire HearGV genome was sequenced, overall nucleotide sequence identity with the XecnGV was found to be 94.8% (Harrison and Popham, 2008). Pairwise Kimura-2-parameter nucleotide distances for lef-8, lef-9, and granulin among XecnGV and previously described isolates ranged from 0.000 to 0.015, but the nucleotide distances between HearGV and these isolates for lef-8 and lef-9 were somewhat greater, ranging from 0.015 to 0.021 and 0.015 to 0.022, respectively. HearGV also possessed a very large genome (169,794 bp) which was extensively co-linear with that of XecnGV (Harrison and Popham, 2008).

The comparison of HearGV and XecnGV genomes also revealed the presence of multiple large indels that distinguish these genomes from each other, indicating that a significant number of genomic rearrangement events had taken place since these two viruses diverged. A total of 16.6 kbp of XecnGV-homologous nucleo-
tide sequence containing 12 ORFs was missing from scattered locations in the HearGV genome, while 8.2 kbp of novel sequence containing nine ORFs was present in the HearGV sequence. A sig-
nificant number of these insertions and deletions either involved bro genes, of which there are 10 in the HearGV genome, or occurred next to bro genes and hr sequences.

The HearGV and XecnGV genomes shared 20 ORFs that were not found among other GVs or NPVs (Harrison and Popham, 2008). These ORFs are uncharacterized, though one (xc168) has a homol-
logue in the genome of the tsetse fly salivary gland hypertrophy virus, and another, xc106, has some sequence similarity to an ORF found in ascoviruses. Several of these ORFs are located near hr3 and hr7, suggesting again a role for hr regions in recombination events leading to acquisition of new genes.

4. Conclusions: 50 genomes are not enough

The increasing availability of genome sequence data from both divergent and closely related baculoviruses has provided a more detailed and comprehensive picture of baculovirus evolution, one that points to an early role for genomic rearrangements in shaping baculovirus genomes as baculoviruses begin to separate into dis-
tinct lineages. Besides insertions and deletions, homologous recombin-
ation among different genotypes of individual baculovi-
rus species has been reported in both field isolates and in cell cul-
ture (Crozier and Ribeiro, 1992; Smith and Crook, 1988; 1993). Recombination between different clade 1a viruses resulting in mosa-
ic genomes has been shown to occur in cell culture (Crozier et al., 1994; Maeda et al., 1993; Summers et al., 1980); and restric-
tion endonuclease mapping of RoMNPV plaque isolates has sug-
gested instances of recombination between RoMNPV and AcMNPV occurring in natural populations (Smith and Summers, 1980). The comparison of clade 1a virus genome sequences pro-
vides more evidence that homologous recombination and allelic replacement occur between different baculoviruses in the field.
Furthermore, the instances of allelic replacement occurring with the polh and ie-2 genes in this group indicate that homologous recombination among baculovirus genomes does not necessarily require an extensive degree of sequence similarity and can occur between distantly related baculoviruses. Finally, comparison of the HearGV and XecnGV genome sequences suggests that an extensive degree of rearrangement can take place as a baculovirus starts to split into different lineages. As more closely related genome sequences become available for analysis, it will be interesting to see if there is a correlation between numbers of homologous repeats and bro genes and the number of rearrangement events observed between virus isolates.

Baculoviruses struggle against evolving host defenses and can encounter altered environments in the form of a genetically different host population (for example, one that may be resistant to virus infection; Asser-Kaiser et al., 2007) or a population of a different host species altogether. Mutations in the form of substitutions can be used to adapt to changing environments, and high mutation rates potentially can be advantageous in a non-static environment (Duffy et al., 2008). However, as with any genetic system, most mutations that occur in virus genomes have a deleterious effect. Hence, one would expect natural selection to favor an upper limit on mutation rates, especially for viruses with large genomes such as baculoviruses (Duffy et al., 2008).

Recombination may serve as a mechanism for baculoviruses to rapidly obtain the variation required for survival in a non-static environment without the potential loss of viability that may occur with a high mutation rate. Comparisons of AcMNPV-like and XecnGV-like genome sequences have revealed a relatively high degree of genomic rearrangement among these viruses. Is there a correlation between the degree of rearrangement observed with these genomes and the large numbers of different hosts from which these viruses have been isolated? Sequencing of more genomes from these clades and from other groups of baculoviruses will be required to explore this possibility.

Acknowledgements

I wish to thank the scientists who have assisted me with my research on baculovirus genomes, including my collaborators Dr. Bryony Bonning (Dept. of Entomology, Iowa State University), Dr. Holly Popham (USDA-ARS-BCIRL, Missouri, Columbia), and Dr. Dwight Lynn (recently retired from the USDA-ARS).

References


Rachiplusia ou (Guenee) (Lepidoptera: NOCTUIDAE). J. Invertebr. Pathol. 8, 21, 167–177.


