



Towards a molecular identification and classification system of lepidopteran-specific baculoviruses

Martin Lange,^a Hualin Wang,^{a,b} Hu Zhihong,^b and Johannes A. Jehle^{a,*}

^aLaboratory for Biotechnological Crop Protection, Department of Phytopathology, Agricultural Service Center Palatinate, 67435 Neustadt an der Weinstraße, Germany

^bJoint Laboratory of Invertebrate Pathology and Key Laboratory of Molecular Virology, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, People's Republic of China

Received 4 March 2004; returned to author for revision 1 April 2004; accepted 15 April 2004

Abstract

Virus genomics provides novel approaches for virus identification and classification. Based on the comparative analyses of sequenced lepidopteran-specific baculovirus genomes, degenerate oligonucleotides were developed that allow the specific amplification of several regions of the genome using polymerase chain reaction (PCR) followed by DNA sequencing. The DNA sequences within the coding regions of three highly conserved genes, namely *polyhedrin/granulin* (*polh/gran*), *late expression factor 8* (*lef-8*), and *late expression factor 9* (*lef-9*), were targeted for amplification. The oligonucleotides were tested on viral DNAs isolated from historical field samples, and amplification products were generated from 12 isolated nucleopolyhedrovirus (NPV) and 8 granulovirus (GV) DNAs. The PCR products were cloned or directly sequenced, and phylogenetic trees were inferred from individual and combined data sets of these three genes and compared to a phylogeny, which includes 22 baculoviruses using a combined data set of 30 core genes. This method allows a fast and reliable detection and identification of lepidopteran-specific NPVs and GVs. Furthermore, a strong correlation of the base composition of these three genome areas with that of the complete virus genome was observed and used to predict the base composition of uncharacterized baculovirus genomes. These analyses suggested that GVs have a significantly higher AT content than NPVs.

© 2004 Elsevier Inc. All rights reserved.

Keywords: *Baculoviridae*; Nucleopolyhedrovirus; Granulovirus; Virus identification; Molecular phylogeny; Polymerase chain reaction

Introduction

Baculoviruses are arthropod-specific, enveloped, rod-shaped viruses with a circular double-stranded DNA genome. The infection of baculoviruses is characterized by a biphasic replication cycle during which two virion phenotypes are produced: (i) the budded virus (BV), which initiates secondary infections and spreads the infection throughout the host, and (ii) the occlusion-derived virus (ODV). The ODVs released from the diseased insects spread the infection in the host population by initiating primary infection of midgut epithelial cells (Federici, 1997). The family *Baculoviridae* is

presently classified into two genera, nucleopolyhedroviruses (NPVs) and granuloviruses (GVs) (Blissard et al., 2000). However, phylogenetic analyses based on gene and genome sequences suggested at least a third genus comprising the dipteran-specific *Culex nigripalpus* (Cuni) NPV (CuniNPV) (Afonso et al., 2001; Moser et al., 2001). On the basis of single gene phylogenies, the lepidopteran-specific NPVs have been divided into two groups, group I NPVs and group II NPVs (Bulach et al., 1999; Zanotto et al., 1993). This separation is also correlated with the distribution of two different envelope fusion proteins that are utilized by the BV for receptor-mediated endocytosis (Pearson and Rohrmann, 2002) and was recently corroborated by gene content, gene order, and whole genome phylogenies of 9 and 13 baculovirus genomes (Herniou et al., 2001, 2003).

About 600 baculoviruses were described to infect insects that belong mostly to the insect orders Lepidoptera, Hymenoptera, and Diptera (Martignoni and Iwai, 1981). About

* Corresponding author. Laboratory for Biotechnological Crop Protection, Department of Phytopathology, Agricultural Service Center Palatinate, Breitenweg 71, 67435, Neustadt an der Weinstraße, Germany. Fax: +49-6321-671222.

E-mail address: Johannes.Jehle@dlr.rlp.de (J.A. Jehle).

90% were reported from 34 different families of Lepidoptera. Considering the diversity of Lepidoptera, it was suggested that there might exist thousands of different baculovirus species in nature (Federici, 1997). Despite this overwhelming diversity, only about 40 of them have been studied in more detail, and to date, the sequences of about 20 baculovirus genomes have been published (Table 1).

Baculoviruses are identified and named according to the insect host species from which they were first isolated. Associations of the same virus with different hosts resulted in the double naming of identical viruses, for example, *Rachiplusia ou* NPV and *Anagrapha falcifera* NPV were shown to be the same virus (Harrison and Bonning, 1999). The number of ambiguous namings among the 500 presently described lepidopteran-specific baculoviruses is unknown. Thus, the real diversity of baculoviruses is unclear due to the lack of a reliable virus identification system. Molecular methods such as DNA restriction endonuclease analyses, gene cloning and sequencing, and molecular phylogeny provided useful tools for gene and genome characterization. In general, however, these methods first require the propagation of the virus in a particular host species, which can be very laborious and time consuming. For AcMNPV and a few other NPVs, it was previously shown that polymerase chain reaction (PCR) can be successfully applied for virus identification (de Moraes and Maruniak, 1997). This method, however, was restricted to a few NPVs and a single genome locus of the *polyhedrin*

gene. To explore the diversity of baculoviruses and to provide a fast and universal identification tool, we examined the genome sequence information of completely sequenced lepidopteran-specific NPVs and GV, and developed degenerate primers of highly conserved gene sequences, which can be amplified by polymerase chain reaction (PCR) and subsequently sequenced. These primers target the *polyhedrin/granulin* gene (*polh/gran*), *late expression factor 8* (*lef-8*), and *late expression factor 9* (*lef-9*) genes that appear to be universal for lepidopteran-specific NPVs and GVs.

Results

Design of degenerate PCR primers

DNA sequence alignments of all conserved genes from published lepidopteran baculovirus genomes (Table 1) were generated and analyzed for highly conserved DNA sequence patterns that could be used as targets for degenerate primers. Initially, more than 15 candidate primer pairs were evaluated in silico and in PCR test reactions for their specificity and amplification performance. Finally, three sets of primers that specifically target the *polh/gran*, *lef-8*, and *lef-9* genes were selected (Table 2). A sample of 20 uncharacterized baculovirus isolates, including 12 nucleopolyhedroviruses (NPVs) and 8 granuloviruses (GVs), on which these primers were successfully used, is given in Table 3.

Table 1
Characteristics of fully sequenced baculovirus genomes

Virus	Abbreviation	Genome size (bp)	Calculated AT content ^a (%)	Modelled AT content ^b (%)	Genbank accession no.	Reference
<i>Autographa californica</i> MNPV	AcMNPV	133,894	59.3	59.6	L22858	Ayres et al., (1994)
<i>Rachiplusia ou</i> MNPV	RoMNPV	131,526	60.9	61.6	NC_004323	Harrison and Bonning (2003)
<i>Bombyx mori</i> NPV	BmNPV	128,413	59.6	60.3	L33180	Gomi et al., (1999)
<i>Epiphyas postvittana</i> NPV	EppoNPV	118,584	59.3	59.0	NC_003083	Hyink et al., (2002)
<i>Orgyia pseudotsugata</i> MNPV	OpMNPV	131,995	44.9	46.5	U75930	Ahrens et al., (1997)
<i>Choristoneura fumiferana</i> MNPV	CfMNPV	129,609	49.9	51.8	NC_004778	–
<i>Adoxophyes honmai</i> NPV	AdhoNPV	113,220	64.4	64.0	NC_004690	Nakai et al., (2003)
<i>Helicoverpa armigera</i> NPV (G4)	HearNPV	131,403	61.0	62.1	NC_002654	Chen et al., (2001)
<i>Helicoverpa armigera</i> NPV (C1)	HearNPV	130,760	61.1	62.2	NC_003094	–
<i>Helicoverpa zea</i> SNP	HzSNPV	130,869	60.9	61.8	NC_003349	Chen et al., (2002)
<i>Mamestra configurata</i> NPV (A)	MacoNPV (A)	155,060	58.3	55.5	NC_003529	Li et al., (2002a)
<i>Mamestra configurata</i> NPV (B)	MacoNPV (B)	158,482	60.0	57.6	NC_004117	Li et al., (2002b)
<i>Lymantria dispar</i> MNPV	LdMNPV	161,046	42.5	43.2	AF081810	Kuzio et al., (1999)
<i>Spodoptera exigua</i> MNPV	SeMNPV	135,611	56.2	54.3	AF169823	Ijkel et al., (1999)
<i>Spodoptera litura</i> NPV	SpltNPV	139,342	57.2	54.7	AF325155	Pang et al., (2001)
<i>Cryptophlebia leucotreta</i> GV	CrleGV	110,907	67.6	68.3	AY229987	Lange and Jehle (2003)
<i>Cydia pomonella</i> GV	CpGV	123,500	54.7	55.5	NC_002816	Luque et al., (2001)
<i>Phthorimea operculella</i> GV	PhopGV	119,217	64.3	61.7	NC_004062	–
<i>Plutella xylostella</i> GV	PlxyGV	100,999	59.3	61.1	NC_002593	Hashimoto et al., (2000)
<i>Adoxophyes orana</i> GV	AdorGV	99,657	65.5	65.1	NC_005038	Wormleaton et al., (2003)
<i>Xestia c-nigrum</i> GV	XecnGV	178,733	59.3	60.7	NC_002331	Hayakawa et al., (1999)
<i>Culex nigripalpus</i> NPV	CuniNPV	108,252	49.1	– ^c	NC_003084	Afonso et al., (2001)

^a AT content calculated from the complete genome.

^b AT content modelled using the regression formula “%AT = 1.0347 × (weighted %AT of *polh/gran*, *lef-8*, and *lef-9*) + 2.913” (compare Fig. 3D).

^c Not calculated due to the lack of a *polyhedrin* sequence.

Table 2

Degenerate oligonucleotide primer sequences used for PCR

Target gene	Primer name	AcMNPV genome position	T _m (°C)	Sequence ^{a,b}
<i>Polh/gran</i>	prPH-1	42075–42088	38–54	<u>TG</u> TAAAACGACGGCCAGTNRNCGARGAYCCNTT
	prPH-2	41373–41389	38–52	<u>CAG</u> GAAACAGCTATGACCDGGNGCRAAYTCYTT
<i>Lef-8</i>	prL8-1	49748–49763	38–47	<u>CAG</u> GAAACAGCTATGACCCAYGGHGARGATGAC
	prL8-2	50027–50043	50–60	<u>CAG</u> GAAACAGCTATGACCAYRTAS ₁ GGRTCYTCS ₂ GC
<i>Lef-9</i>	prL9-1	4684–4698	40–53	<u>CAG</u> GAAACAGCTATGACCAARAAYGGITAYGCB ₃ G
	prL9-2	5210–5224	47–57	<u>TG</u> TAAAACGACGGCCAGTTT ₄ TDCRCRCRCARTC

^a (B = C, G, or T; D = A, G, or T; H = A, C, or T; I = Inosin; N = C, A, T, or G; R = A or G; S = C or G; Y = C or T).

^b Underlined nucleotides indicate standard sequencing primers (–21) M13 forward and (–29) M13 reverse. Nucleotides with subscripted numbers indicate mismatches to the target sequence of the baculoviruses listed in Table 1: (S₁ = T in CrleGV, S₂ = T in AdhoNPV, B₃ = A in AdhoNPV, G₄ = A in PhopGV and HzSNPV). The given T_m range is related only to the degenerate (non-underlined) part of the primer sequence.

The degenerate primers prPH-1 and prPH-2 match their target regions within the *polh/gran* genes in all 21 lepidopteran-specific NPVs and GVs sequenced so far (Table 1). The sizes of amplified PCR products of 20 baculoviruses using these primers ranged from 507 to 510 nucleotides and represented the major portion of the *polh* gene (Table 3). Only in a very few cases did these fragments reveal any sequence ambiguities. For *Cnephasia longana* (Cnlo) GV (AY519225), three ambiguities were observed, none of which led to a change in the amino acid. For *Coloradia pandora* (Copa) NPV (AY519228), a single nucleotide uncertainty resulted in two possible amino acid residues (isoleucine and phenylalanine). Because all other known *polh* sequences encode an isoleucine residue at this position, it is very likely that this ambiguity was caused by a sequencing error. Therefore, only the isoleucine residue was considered in the phylogenetic analyses. The *polh/gran* DNA sequences were identical for *Malacosoma californicum* (Maca) NPV (AY519237) and *Malacosoma* sp. NPV (AY519240), and for *Hoplodrina ambigua* (Hoam) GV (AY519231) and *Autographa gamma* (Auga) GV (AY519213), respectively. The rates of synonymous and non-synonymous nucleotide substitutions for selected NPVs and GVs are given in Table 4.

The oligonucleotides prL8-1 and prL8-2 are *lef-8* specific. Primer prL8-1 matched its target sequence within the *lef-8* gene in all lepidopteran-specific baculoviruses, whereas prL8-2 revealed one mismatch to the *lef-8* sequences of the recently sequenced *Cryptophlebia leucotreta* (Crle) GV and *Adoxophyes honmai* (Adho) NPV (Table 2). However, the primers prL8-1 and prL8-2 have been used to amplify partial *lef-8* sequences of more than 100 different lepidopteran baculovirus isolates including CrleGV, suggesting that these mismatches are not crucial for primer specificity (data not shown). The *lef-8* specific primers produced PCR fragments at annealing temperatures ranging from 38 to 48 °C. Thus, due to the degeneracy of these primers, an optimization of the PCR conditions was crucial for the amplification of target genes. The size of the amplification products obtained by using prL8-1 and prL8-2 ranged from 681 nucleotides for

Trichoplusia ni (Tn) GV (AY519202) to 771 nucleotides for *Pieris rapae* (Pira) GV (AY519253).

The oligonucleotides prL9-1 and prL9-2 are *lef-9* specific. Recently published genome sequences revealed a few mismatches that were not covered by the degeneracy of prL9-1 and prL9-2 (Table 2). Primer prL9-1 had a single mismatch near its 3' terminus to the AdhoNPV *lef-9* sequence, whereas prL9-2 had a single mismatch to its target of *Phthorimea operculella* (Phop) GV and HzSNPV, respectively. However, the mismatching nucleotides of prL9-2 are near the 5' end of the primer. Because the amplification of the partial *lef-9* gene of BufuNPV, which was identical to the published *lef-9* gene of HzSNPV (NC_003349), was not impaired, it is most likely that this mismatch is not crucial for primer binding and specificity. So far, the *lef-9* primers have been used to amplify partial *lef-9* genes of more than 100 different virus isolates, underscoring their broad specificity for lepidopteran baculoviruses (Lange and Jehle, unpublished). Amplification products were obtained at annealing temperatures ranging from 45 to 54 °C. Occasionally, an amplification reaction yielded products of different sizes, of which one disappeared by increasing the temperature. As described above for the *lef-8* primer pairs, an optimization of the PCR conditions was also sometimes needed for the *lef-9* specific primers depending on the virus isolate. PrL9-1 and prL9-2 generated fragments of 258 nucleotides for PiraGV (AY519254) and *Amelia pallorana* (Ampa) GV (AY519209) to 273 nucleotides for HoamGV and AugaGV (Table 3). The DNA sequences obtained for MacaNPV and *Malacosoma* sp. NPV were identical to each other, as were those of BufuNPV and HzSNPV, and those of AdorGV, and the published AdorGV *lef-9* sequence.

Phylogenetic information of concatenated partial *lef-8*, *lef-9*, and *polh* genes

The phylogenetic relationship of 21 baculoviruses was reconstructed from deduced amino acid sequences of 30 conserved baculovirus core genes that are present in all sequenced genomes of lepidopteran-specific baculoviruses (Herniou et al., 2003, Lange and Jehle, 2003). The concatenated alignment contained 16273 characters of

Table 3

Amplified and sequenced *polh/gran*, *lef-8*, and *lef-9* partial genes from different lepidopteran-specific NPVs and GVs

Virus	Abbreviation	Host family	CS ^a	Partial gene sequences; Genbank accession no. ^b Length (nt)/%AT			Predicted genome AT content ^c (%)
				<i>polh/gran</i>	<i>lef-8</i>	<i>lef-9</i>	
<i>Phryganidia californica</i> NPV	PhcaNPV	Dioptriidae	B	AY519249 507/54.4 ^d	AY519250 687/54.6	AY519251 261/49.0	58.3
<i>Coloradia pandora</i> NPV	CopaNPV	Saturniidae	B	AY519228 510/59.2	AY519229 696/63.8	AY519230 246/53.7 ^c	65.5
<i>Junonia coenia</i> NPV	JucoNPV	Nymphalidae	B	Ay519234 510/50.4	AY519235 684/60.2	AY519236 261/47.9	59.4
<i>Bombyx mori</i> NPV	BmNPV	Bombycidae	B	AY519216 510/51.8	AY519217 684/61.3	AY519218 261/48.3	60.4
<i>Aporia crataegi</i> NPV	ApcrNPV	Pieridae	D	AY519210 510/57.3	AY519211 687/63.2	AY519212 261/52.1 ^d	64.1
<i>Pterolocera amplicornis</i> NPV	PtamNPV	Anthelidae	B	AY519255 510/55.9	AY519256 684/62.9	AY519257 261/54.0	63.8
<i>Boamria bistortata</i> NPV	BobNPV	Geometridae	D	AY519219 494/50.4 ^c	AY519220 729/59.7	AY519221 261/51.0	59.9
<i>Malacosoma neustria</i> NPV	ManeNPV	Lasiocampidae	D	AY519243 510/54.5	AY519244 699/55.8	AY519245 261/49.8	59.1
<i>Malacosoma sp.</i> NPV	—	Lasiocampidae	B	AY519240 510/54.5	AY519241 699/54.0	AY519242 261/46.0	57.5
<i>Malacosoma californicum</i> NPV	MacaNPV	Lasiocampidae	B	AY519237 510/54.5	AY519238 699/53.8	AY519239 261/46.0	57.4
<i>Agrotis ipsilon</i> NPV	AgipNPV	Noctuidae	I	AY519204 510/48.0	AY519205 702/42.3	AY519206 261/37.6	47.9
<i>Busseola fusca</i> NPV	BufuNPV	Noctuidae	D	AY519222 510/58.8	AY519223 742/57.0 ^c	AY519224 261/52.9	61.8
<i>Trichoplusia ni</i> GV	TnGV	Noctuidae	D	AY519201 510/49.4	AY519202 681/64.2	AY519203 267/54.3	62.1
<i>Hoplodrina ambigua</i> GV	HoamGV	Noctuidae	D	AY519231 510/48.2	AY519232 681/62.4	AY519233 273/56.0	61.2
<i>Adoxophyes orana</i> GV	AdorGV	Tortricidae	D	AY519258 510/52.8	AY519259 655/67.9 ^c	AY519260 261/55.9	65.1
<i>Autographa gamma</i> GV	AugaGV	Noctuidae	D	AY519213 510/48.2	AY519214 681/62.1	AY519215 273/56.4	61.1
<i>Cnephasia longana</i> GV	CnloGV	Tortricidae	D	AY519225 490/54.5 ^{d,c}	AY519226 716/69.7 ^d	AY519227 261/60.5	68.1
<i>Pieris rapae</i> GV	PiraGV	Pieridae	B	AY519252 507/56.0	AY519253 771/69.4	AY519254 258/60.9	68.7
<i>Amelia pallorana</i> GV	AmpaGV	Tortricidae	B	AY519207 510/58.6	AY519208 723/74.1	AY519209 258/67.8	73.1
<i>Pandemis limitata</i> GV	PaliGV	Tortricidae	B	AY519246 510/55.1	AY519247 723/67.0	AY519248 261/62.1	67.1

Given are the Genbank accession numbers along with the nucleotide lengths (nt) and the AT contents (%AT) of the different amplification products of the indicated viruses. The given genome AT content (%) is predicted from the regression between the genomic AT contents of complete baculovirus genomes and the combined AT contents of the partial gene sequences.

^a CS = collection Source (B = Berkeley, US; D = Darmstadt, Germany; I = Iowa State University, US).

^b The sequences were deposited in Genbank under the given accession numbers.

^c Calculated by the regression formula “%AT = 1.0347 (weighted %AT of *polh/gran*, *lef-8*, and *lef-9*) + 2.913” (compare Fig. 3D).

^d AT content may vary by 0.6% due to ambiguous sequence positions.

^e Ambiguous nucleotide positions located at the 5' terminus or 3' terminus were not included.

which 2728 positions were constant and 10844 positions were parsimony informative. Fig. 1A shows a parsimony tree inferred by a heuristic search including the bootstrap support from a maximum parsimony (MP), neighbor joining distance (NJ), and minimum evolution (ME) analysis. The distantly related dipteran-specific CuniNPV was included as the outgroup. Within the lepidopteran-specific baculoviruses, three clusters with high bootstrap support

were distinguished corresponding to group I and group II NPVs and the GVs as hypothesized by Zanutto et al. (1993) and Bulach et al. (1999). The branching order between all taxa belonging to NPV group I was supported 100% in all three bootstrap analyses. The taxa of group I NPVs were further subdivided into two distinct clades. In contrast, the clade containing the group II NPVs showed lower bootstrap support, and conflicting tree topologies in

Table 4

Rates of synonymous (Sd) and non-synonymous (Sn) nucleotide substitutions in partial *polh*, *lef-8*, and *lef-9* genes of baculoviruses listed in Tables 1 and 3

Compared viruses	<i>polh</i>		<i>lef-8</i>		<i>lef-9</i>	
	Sd	Sn	Sd	Sn	Sd	Sn
AcMNPV (L22858) vs. RoMNPV (NC_004323)	73.33	37.67	22	7	0	1
AcMNPV (L22858) vs. BmNPV (L33180)	74.83	41.17	15	10	7	1
AcMNPV (L22858) vs. BmNPV (AY519216–AY519218)	76	42	16	11	5	1
AcMNPV (L22858) vs. JucoNPV (AY519234–AY519236)	1	0	0	0	5	1
RoMNPV (NC_004323) vs. BmNPV (L33180)	26	9	29	17	7	0
RoMNPV (NC_004323) vs. BmNPV (AY519216–AY519218)	27	9	30	18	5	0
BmNPV (L33180) vs. BmNPV (AY519216–AY519218)	2	4	5	1	2	0
HsSNPV (NC_003349) vs. BfuvNPV (AY519222–AY519224)	4	0	0	2	0	0
HsSNPV (NC_003349) vs. HearNPV (NC_002654)	4	0	5	3	1	0
BfuvNPV (AY519222–AY519224) vs. HearNPV (NC_002654)	6	0	5	3	1	0
XecnGV (NC_002331) vs. HoamGV (AY519231–AY519233)	1	0	7	1	2	0
XecnGV (NC_002331) vs. AugaGV (AY519213–AY519215)	1	0	8	2	3	0
HoamGV (AY519231–AY519233) vs. AugaGV (AY519213–AY519215)	0	0	1	1	1	0
TnGV (AY519201–AY519203) vs. XecnGV (NC_002331)	53	1	82.17	28.83	37.50	7.50
TnGV (AY519201–AY519203) vs. HoamGV (AY519231–AY519233)	54	1	84.17	27.83	38.50	7.50
TnGV (AY519201–AY519203) vs. AugaGV (AY519213–AY519215)	54	1	85.17	28.83	37.50	7.50
ManeNPV (AY519243–AY519245) vs. Malacosoma sp. NPV (AY519240–AY519242)	41	1	51	18	15.50	2.50
ManeNPV (AY519243–AY519245) vs. MacaNPV (AY519237–AY519239)	41	1	50	18	15.50	2.50
Malacosoma sp. NPV (AY519240–AY519242) vs. MacaNPV (AY519237–AY519239)	0	0	1	0	0	0
AdorGV (NC_005038) vs. AdorGV (AY519258–AY519260)	1	0	0	0	0	0

the branching order of AdhoNPV and *Lymantria dispar* (Ld) MNPV were observed. Whereas the MP and ME analyses could not resolve the branching order of AdhoNPV and LdMNPV, the NJ distance consensus tree placed LdMNPV at the base of all taxa belonging to group II NPVs, followed by AdhoNPV (75% bootstrap support, data not shown). A similar situation was observed for the branching order of the GVs. The MP consensus tree placed XecnGV at the base of all GVs, followed by *Plutella xylostella* (Plxy) GV (Fig. 1A), whereas in the ME and NJ, distance consensus trees XecnGV and PlxyGV clustered together with high bootstrap support (data not shown).

To validate the phylogenetic information derived from the PCR amplification products of *polh/gran*, *lef-8*, and *lef-9*, we compared the phylogeny of the putative PCR amplification products of the 22 sequenced genomes to that of their 30-core gene tree (Figs. 1A and B). The topology of the parsimony consensus tree of the partial sequences was compatible with the 30-core gene tree and resulted in the best bootstrap values of all analyses where single partial gene trees were used (data not shown). Thus, the sequence information obtained from these partial genes successfully mirrored the phylogeny based on 30-core genes, which is presently considered to represent the “true” baculovirus phylogeny.

Phylogenetic analyses of uncharacterized baculoviruses

Table 3 summarizes the sequence information obtained for *polh/gran*, *lef-8*, and *lef-9* from 20 uncharacterized baculovirus isolates. The phylogenetic tree based on concatenated partial *polh/gran*, *lef-8*, and *lef-9* amino acid sequences from 42 baculoviruses (listed in Tables 1 and 3)

is given in Fig. 2. These analyses provided a significant support for the GVs and group I NPVs, as well as a less stable branching support for group II NPVs. The tree topology is completely compatible with the 30-core gene tree of Fig. 1A and resulted in the best bootstrap values of all analyses where partial gene trees were used (data not shown).

Out of the 20 analyzed baculoviruses listed in Table 3, 8 viruses clustered within the GVs (HoamGV, AugaGV, TnGV, CnloGV, PaliGV, AmpaGV, AdorGV, and PiraGV), 6 viruses clustered with group I NPVs (PhcaNPV, CopaNPV, PtamNPV, BmNPV, ApcrNPV, and JucoNPV), and 6 clustered with group II NPVs (AgipNPV, BfuvNPV, MacaNPV, Malacosoma sp. NPV, ManeNPV, and BobiNPV). With the exception of the *polh* gene of JucoNPV, which differs in only one synonymous nucleotide substitution to that of AcMNPV (Table 4), all single gene trees that were inferred from partial *lef-8*, *lef-9*, and *polh/gran* amino acid sequences resulted in a consistent grouping of the viruses (data not shown). This indicates that any of the primer pairs could be used for identification and group allocation of the analyzed baculoviruses. The phylogenetic analyses (Fig. 2A) and the nucleotide substitution rates (Table 4) indicated that AcMNPV is virtually the same virus as the analyzed JucoNPV isolate. The same holds true for XecnGV, HoamGV, and AugaGV, as well as for HsSNPV and BfuvNPV.

Prediction of genome AT content

The AT contents of the completely sequenced baculovirus genomes were compared to the AT content of the complete open reading frames of *polh/gran*, *lef-8*, and

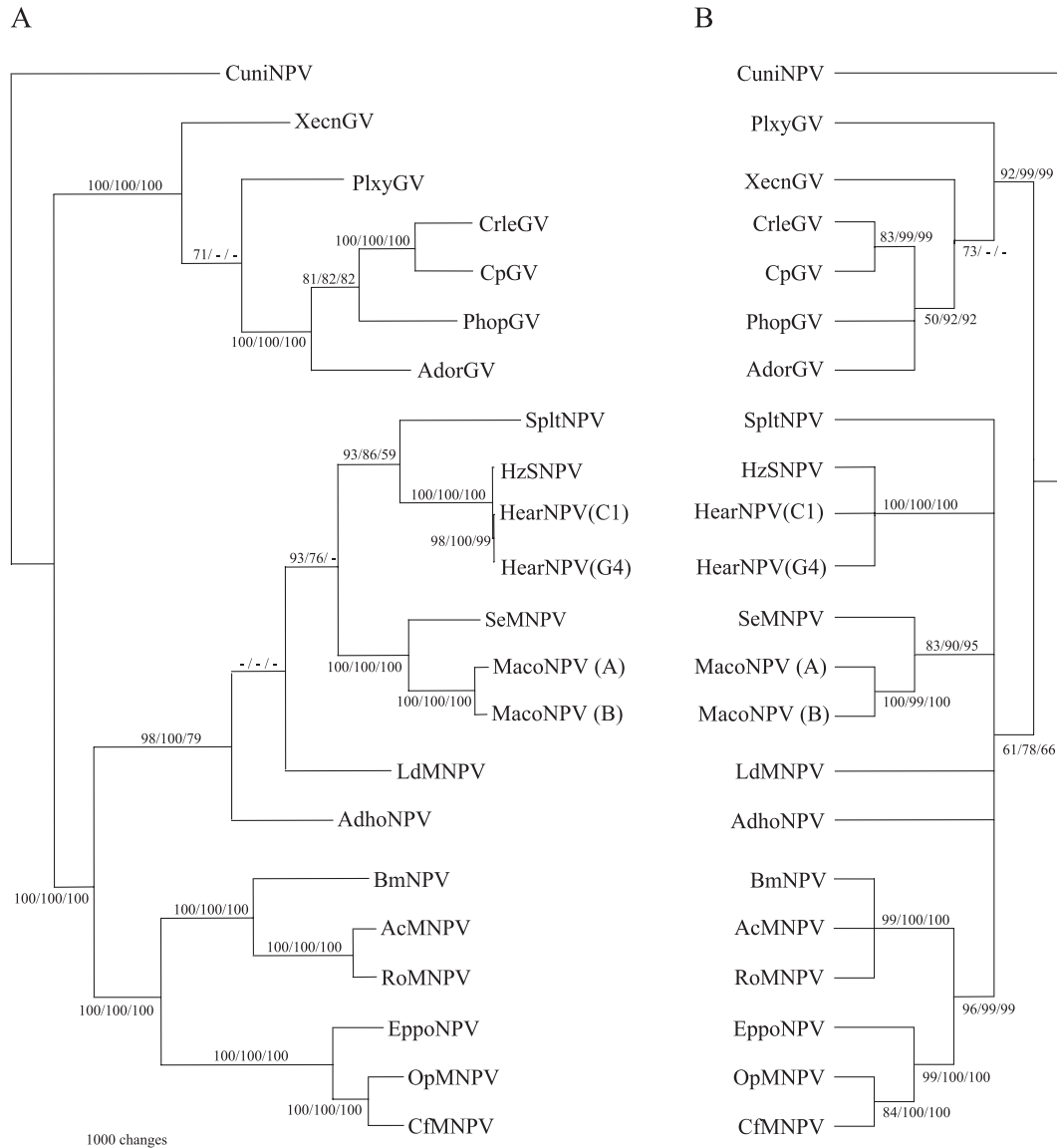


Fig. 1. (A) Baculovirus phylogram based on amino acid sequences of 30 individually aligned and concatenated baculovirus core genes. Given is a maximum parsimony tree (tree length = 72453, consistency index = 0.7556). Numbers above or below the nodes indicate the bootstrap values obtained for maximum parsimony (MP), neighbor joining (NJ) distance, and minimum evolution (ME) analyses, respectively (=MP/NJ/ME). (B) Maximum parsimony consensus tree (cladogram) based on concatenated amino acid sequences of the partial *polh/gran*, *lef-8*, and *lef-9* genes from published baculovirus genomes given in (A). Numbers at the nodes indicate the bootstrap values from MP, NJ distance, and ME analyses, respectively (=MP/NJ/ME). The dipteran-specific CuniNPV was used as outgroup in both analyses.

lef-9 as well as the weighted means of these sequences. Alternatively, only the partial gene segments that are amplified using the developed degenerate oligonucleotides (Table 2) were considered (Figs. 3A–D). As shown by the calculated regression lines and regression coefficients, there was only a poor correlation ($R^2 = 0.3994$ and 0.4181) between the genome and *polh/gran* AT contents (Fig. 3A). Although the *lef-8* and *lef-9* sequences showed a better correlation with the genome AT content (Figs. 3B and C), the best regression coefficients were obtained when the combined set of sequences were weighted with their length as a factor. Using the regression line $\%AT = 1.0347 \times (\%AT_{polh/gran} \times \text{length}_{polh/gran}$

$+ \%AT_{lef-8} \times \text{length}_{lef-8} + \%AT_{lef-9} \times \text{length}_{lef-9}) / (\text{length}_{polh/gran} + \text{length}_{lef-8} + \text{length}_{lef-9}) + 2.913$, the regression coefficients R^2 were 0.9655 for the complete open reading frames and 0.938 for the partial sequences (Fig. 3D). The predicted genome AT contents, based on this regression of the weighted means of the partial *polh/gran*, *lef-8*, and *lef-9*, differed generally not more than 1–2% from the actual AT content of 21 completely sequenced lepidopteran-specific baculovirus genomes (Table 1). This strong correlation between the AT contents of the whole genomes and the partial genes allowed for the making of reliable predictions of the AT content of the genomes of the viruses considered in this study. As

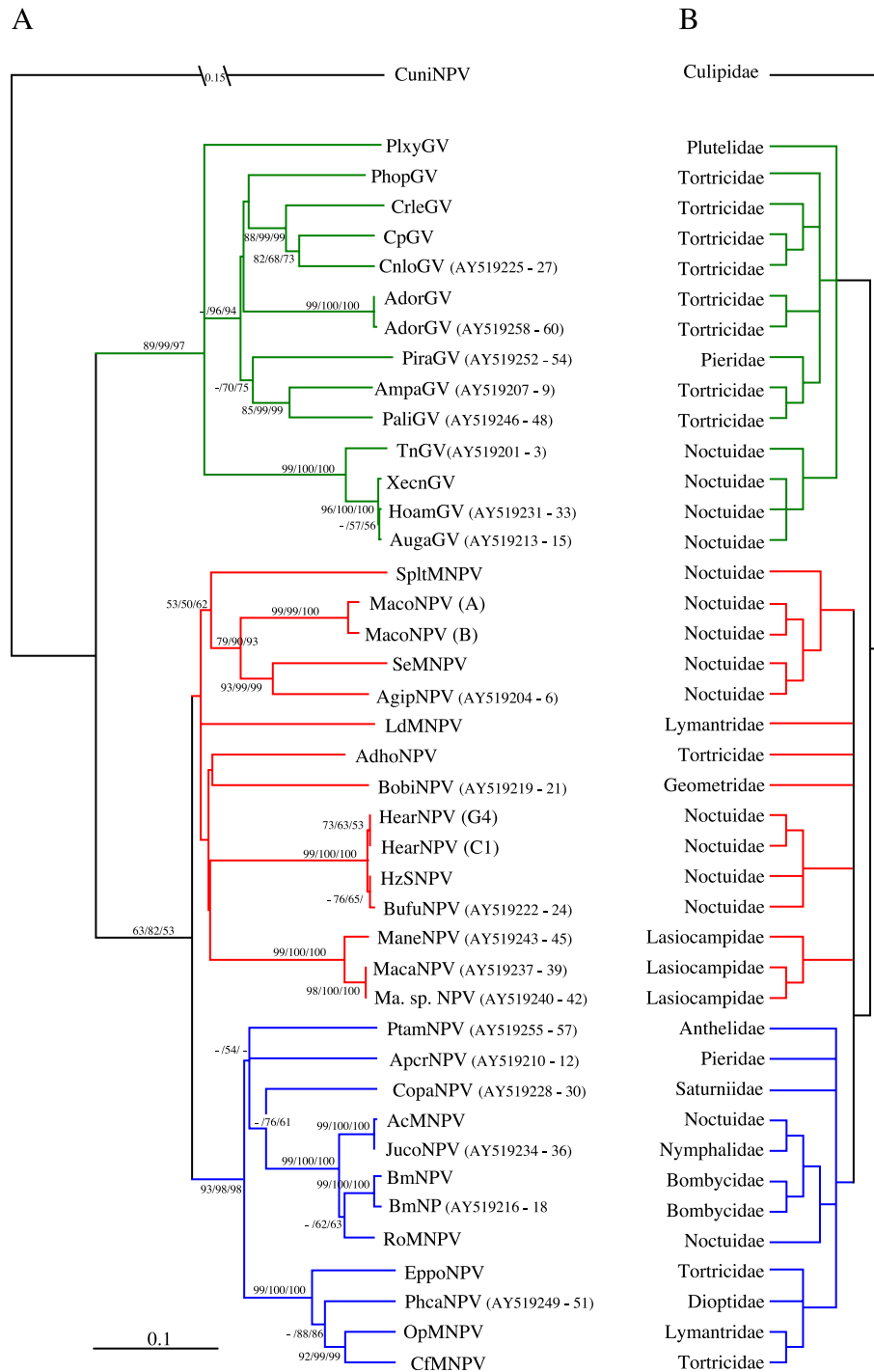


Fig. 2. (A) Neighbor joining (NJ) distance tree based on concatenated amino acid sequences of the partial *polh/gran*, *lef-8*, and *lef-9* genes (for accession numbers, see Tables 1 and 3). Numbers at the nodes indicate the bootstrap values from a maximum parsimony (MP), neighbor joining (NJ) distance, and minimum evolution (ME) analyses, respectively (=MP/NJ/ME). CuniNPV was used as outgroup. The branch colors indicate the group affiliation (blue denotes group I NPVs, red denotes group II NPVs, and green denotes GVs). (B) MP consensus tree based on concatenated amino acid sequences of the partial *polh/gran*, *lef-8*, and *lef-9* genes (for further description and MP bootstrap values, see A). The host family of the respective baculovirus is given at the leaf node.

shown in Table 3, the predictions of the AT contents varied between 47.9% for AgipNPV and 73.1% for AmpaGV. If these predictions are correct, then AmpaGV would have the highest AT content of any known baculovirus.

Discussion

The objective of this study was the development of a fast and high throughput screening method for the identification and classification of lepidopteran baculoviruses. Extensive

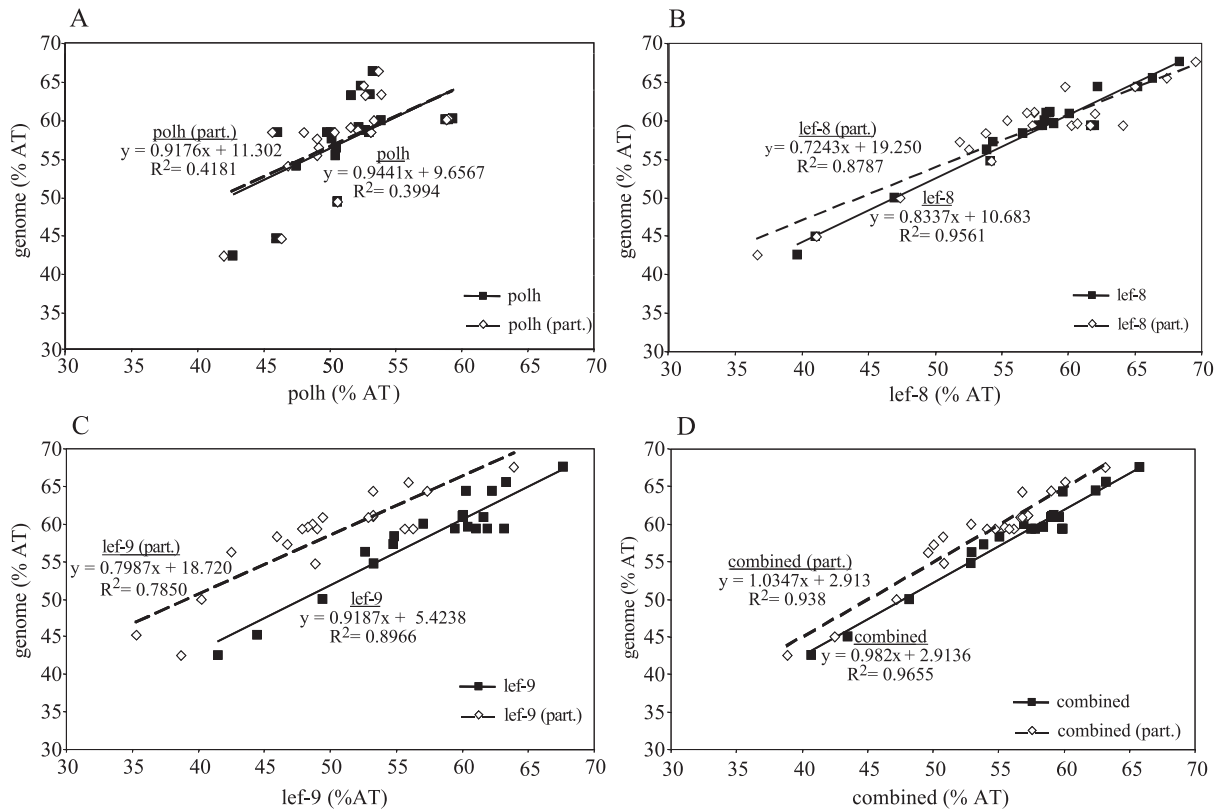


Fig. 3. Correlation between the AT contents of complete genomes and single genes, (A) *polh/gran*, (B) *lef-8*, (C) *lef-9*, and (D) weighted means of *polh/gran*, *lef-8*, and *lef-9*. The data were compiled from all lepidopteran-specific baculoviruses listed in Table 1. The complete gene sequences (solid line) and the partial sequences defined by the positions of the degenerate primers (dashed lines) were used for regression analyses. The formula of the regression lines and the regression coefficients are given in the panels.

comparisons of completely sequenced baculovirus genomes were performed to find genome regions of highly conserved genes that are present in all lepidopteran baculoviruses and can be used as targets for PCR amplification and phylogenetic studies. By analyzing nine baculovirus genomes, Herniou et al. (2001) suggested seven genes that are particularly useful to meet this requirement. These were the homologues to the AcMNPV genes *ac22* (*pif-2*), *ac81*, *ac119* (*pif*), *ac142*, *145*, *lef-8*, and *lef-9*. A further prerequisite for the potential target genes was that they needed highly conserved sequence patterns over a distance of 200–700 nucleotides that allowed for the annealing of degenerate primers. This fragment size ensured that not only short sequence fragments with low information content were avoided, but also allowed for a single sequencing reaction to cover the complete PCR fragments. Finally, standard primer requirements, such as avoidance of hairpin, primer dimer formation, and low specificity to other genome areas, and similar melting temperature for the primer pairs, had to be scrutinized for different extents of sequence degeneration. Out of 15 initially considered primer pairs, three sets targeting the *polh/gran*, *lef-8*, and *lef-9* gene were designed and extensively tested for a wide range of baculoviruses.

The *polh/gran* gene locus is one of the most conserved genes of lepidopteran-specific baculoviruses and was also

the first baculovirus gene that was used for phylogenetic studies (Rohrmann, 1986). Because the tree topology of *polh/gran* did not match the tree topology of concatenated sequences, the *polh/gran* gene locus was discredited for use in phylogenetic analyses in previous studies (Harrison and Bonning, 2003; Herniou et al., 2001). One of the reasons for the discrepancy for *polh* is that the AcMNPV *polh* is most likely a mosaic gene generated by recombination with a group II NPV (Jehle, 2004). Use of the AcMNPV *polh* therefore adds some distortion and instability to the *polh/gran* tree. This could also be observed in the partial *polh/gran* tree (data not shown), where AcMNPV and JucoNPV group within the group II NPVs, whereas the 30-core gene tree (Fig. 1A) clearly indicates that AcMNPV and JucoNPV belong to group I NPVs. For simple molecular identification and combined phylogenetic analyses, *polh/gran* is still very useful, especially because a great number of *polh/gran* gene sequences are available. The other two target genes *lef-8* and *lef-9* encode for the subunits of the baculovirus-encoded RNA polymerase, which initiates transcription from late and very late baculovirus promoters (Guarino et al., 1998). They were previously shown to be highly useful for studying baculovirus phylogeny (Herniou et al., 2001, 2003).

Although the developed primers are highly degenerate, they are still extremely specific for baculovirus DNA. So far,

PCR primers suitable for identifying noctuid-specific NPVs were established by only de Moraes and Maruniak (1997), but their method was far less universal than the approach described in this study, and it pointed only to a single gene, the *polh* locus. Recently, Herniou et al. (2004) presented a phylogenetic study of uncharacterized baculoviruses where *lef-8* and *ac22* (*pif-2*) were used as amplification targets of degenerate PCR primers. Their *lef-8*-specific primers were even evaluated for hymenopteran-specific baculoviruses. But their amplified partial *lef-8* fragment was about 450 nucleotides compared to 684–723 nucleotides obtained in our study.

Virus DNA used for PCR amplification was obtained by complete DNA isolation of virus-infected larvae, indicating that insect DNA as a background neither disturbed the reaction nor resulted in false-positive amplification products. Control experiments where isolated genomic DNA of several uninfected insect larvae was used never resulted in a positive PCR signal (Lange and Jehle, unpublished). All isolates derived from historic field samples, some of which could be traced back to be 40–50 years old (Steinhaus and Marsh, 1962). Considering the nomenclature of these virus samples, we completely depended on the information given on the sample vials or the written records. Many of the samples were even not recognizable as insects. Although it cannot be completely ruled out that some of these historic samples may also contain misidentifications of the host insects, we consider the molecular identification of these samples as the only way to bring more light into the huge amount of baculovirus isolates accumulated during the last 50 years. Using a similar PCR approach with insect specific primers may even provide the opportunity of a retrospective identification of the infected host species from the DNA preparations we made. It is not known if any of these virus samples are still infectious. But nevertheless, as demonstrated, the application of degenerate PCR combined with molecular phylogeny provides an excellent method for fast and reliable baculovirus identification and needs only tiny traces of a sample even if the virus cannot be recovered.

The number of fully sequenced baculovirus genomes is still low compared to the potential number of species present in nature, thus there is always a risk that the developed degenerate primers will not match every potential target. On the other hand, the primers were shown to amplify a very diverse set of lepidopteran-specific baculoviruses including group I and group II NPVs as well as GV. The AT contents of the different *lef-8* and *lef-9* amplification products varied from the low 40% to 74%, indicating that these primers cover a very wide range of sequence variability. So far, between 70 and 100 different baculovirus isolates were successfully analyzed by using these primers, underscoring their broad specificity (Jehle and Lange, unpublished). In addition, it is highly likely that at least one of the primer pairs would amplify their target gene. Therefore, we propose that these primers allow the

amplification and direct sequencing of the target genes from most, if not all, lepidopteran baculoviruses.

The comparative phylogenetic analyses of the *polh/gran*, *lef-8*, and *lef-9* sequences demonstrated that the most reliable results were obtained when the combined set of data was used for tree inference. The trees in Figs. 1B and 2 have no conflicts with the tree derived from 30 concatenated core genes of 22 completely sequenced baculovirus genomes. All single gene trees (with the exception of AcMNPV–JucoNPV *polh*) allowed a classification of the viruses within group I and group II NPVs and GVs. Thus, a single target gene, such as the *polh/gran*, might be in many cases sufficient for a quick identification of an uncharacterized baculovirus isolate. Using the sequence information of all three targets increases the sequence information or allows for the detection of potential conflicts, for example, ambiguous amplification from mixed infected larvae. For safeguarding and further phylogenetic analyses, the combination of the three target genes appears to be preferable.

Sequence comparison of the partial target genes indicated that XecnGV, HoamGV, and AugaGV differ only in a very few nucleotide substitution in these partial genes and are most likely isolates of the same virus, which were apparently isolated from three different hosts. The same holds true for HzSNPV, HearNPV, and BufuNPV as well as JucoNPV and AcMNPV. This was clearly confirmed by all single gene trees and the concatenated gene tree. Apparently, TnGV and XecnGV form a cluster of closely related viruses similar to the AcMNPV–RoMNPV–BmNPV complex. The analyses clearly indicated that in contrast to GVs and group I NPVs, the group II NPVs do not represent a well-defined phylogenetic clade (Fig. 2). This finding is conform with previous analyses of the distribution of group-specific genes that showed that group II NPVs compose a less homogenous group of baculoviruses than GVs and group I NPVs (Lange and Jehle, 2003). It appears that group II NPVs presently subsume all lepidopteran-specific NPVs not belonging to the monophyletic group I NPVs. The distribution of the virus host families (Fig. 2B) so far suggests a strong coevolution between baculoviruses and insect host families. GVs were analyzed from Plutellidae, Tortricidae (which included *Pieris*), and Noctuidae host families and formed separate clades according to their host family specificity. Group II NPVs contained two noctuid-specific clades, that is, a *Heliothis*- and *Spodoptera*-specific virus complex. The picture of virus host coevolution becomes more complex when analyzing those viruses with a broad host range. For example, AcMNPV and BmNPV are known to infect insect species from at least 10 and 7 lepidopteran families, respectively (Gröner, 1986). Studying baculovirus diversity by using many baculovirus isolates will help clarify nomenclatural ambiguities, where different names were given to isolates of the same virus, as was apparently the case for XecnGV, HoamGV, and AugaGV, or HearNPV and BufuNPV, or AcMNPV and JucoNPV. These findings indicate that, for example, XecnGV is most likely also infective for *H. ambi-*

gua and *A. gamma*, and thus makes suggestions for directed host range testing experiments. In contrast, three isolates of “*Galleria mellonella* NPV”, which are currently under investigation, revealed that one isolate is virtually identical with AcMNPV, a second isolate is a group I NPV, whereas the third isolate belongs to group II NPVs (Jehle and Lange, unpublished). In this specific case, the same name was apparently given to three different viruses.

As shown by the regression analyses of AT contents of completely sequenced baculovirus genomes and the three partial gene sequences, the identified targets *polh/gran*, *lef-8*, and *lef-9* can be used to predict the overall AT contents of the viral genomes under investigation. This correlation is remarkable because the AT content of individual genes generally differs from the whole genome by $\pm 10\%$ (Lange and Jehle, 2003; Wormleaton et al., 2003). The *polh* gene appears to be biased to very low AT contents, as was suggested by the steep slope and the poor correlation which both indicate that the AT compositions of many baculovirus *polh* are saturated (Fig. 3A). A bias in the codon composition of a gene could favor high expression levels as was observed for bacteria (Gouy and Gautier, 1982). Considering all known and predicted AT contents in Tables 1 and 3, it is suggested that GVs generally have considerable higher AT contents (average 64%) than group I NPVs (59%) and group II NPVs (58%). It is proposed that the high AT contents of GVs are the result of a not yet elucidated evolutionary constraint rather than the reminiscence to a common high AT ancestor, because the narrowly related CpGV and CrleGV show extreme divergence in base composition (Lange and Jehle, 2003). Knowledge and prediction of AT contents may contribute to the understanding of the evolution and adaptation mechanisms of GVs and NPVs.

According to the present concept of virus classification and taxonomy, a virus species is a polythetic class whose members always have several properties in common, but no single property must be present in all the members of the species (van Regenmoertel, 2000). As a consequence, a single discriminating character, for example, sequence information alone, is never sufficient to define a species. Therefore, a universal identification tool for lepidopteran-specific baculoviruses, as presented in this study, will not replace a proper taxonomic description and classification, but provides a useful molecular tool to aid in the natural classification and nomenclature of baculoviruses by resolving equivocal virus–host associations. To avoid further ambiguous naming of baculoviruses, it is recommended to apply this simple approach for an initial identification of newly isolated viruses.

Materials and methods

Design of degenerate oligonucleotides

Available core genes from fully sequenced lepidopteran baculovirus genomes (Table 1) were downloaded from

Genbank and imported to the BioEdit program (Hall, 1999). Multiple alignments were performed using the ClustalW program (Thompson et al., 1994). Degenerate oligonucleotides were designed from highly conserved genome regions and tested for hairpin and primer–primer interactions by using Lasergene software (DNASTAR, Inc., Madison, WI, USA). Three sets of oligonucleotide primers targeting the *polh/gran*, *lef-8*, and *lef-9* genes were finally selected. Oligonucleotides were synthesized by MWG (Ebersberg, Germany) (Table 2).

DNA extraction from diseased insects

Samples of baculovirus-infected insects were obtained from the University of California, Berkeley (Edward Steinhaus Collection) and the Institute for Biological Control, Federal Biological Research Center for Agriculture and Forestry, Darmstadt, Germany. DNA extractions were performed using the DNeasy Tissue Kit (Qiagen, Hilden, Germany) according to instructions of the manufacturer. To our knowledge, these samples were kept frozen at -20°C , some of them for up to 40–50 years. The samples consisted of partially purified virus suspensions or of the remains of infected larvae. Most of the sample larvae were completely desiccated and were thus not identifiable. For nomenclature of the samples, we had to trust on the records kept at the collecting institutes or solely on the handwritten labels of the collection vials. If possible, only parts of the remains were used for the analysis.

PCR amplification using degenerate primers

Each 50 μl PCR reaction contained 1–100 ng of prepared template DNA, 100 μM of each dATP, dCTP, dGTP, and dTTP, 5 μl of $10\times$ reaction buffer (500 mM Tris–HCl, pH 8.8, 160 mM $(\text{NH}_4)_2\text{SO}_4$, 0.1% Tween 20), 2 mM MgCl_2 , 1 unit Taq DNA polymerase (Axon, Kaiserslautern, Germany), and 0.2 μM of each primer. Amplification reactions were performed using an Eppendorf Master Gradient thermocycler.

For amplification of partial *polh/gran* genes, the primers prPH-1 and prPH-2 that contained the target regions for (–21) M13 forward and (–29) M13 reverse sequencing primers were used (Table 2). The PCR consisted of an initial denaturation step of 95°C for 3 min and 36 cycles of 95°C for 30 s, 70°C for 1 min, 50°C for 1 min, and a final extension step 72°C for 10 min. PCR products of partial *lef-8* genes that were selected for subsequent cloning were obtained using primers prL8-1 and prL8-2 without attached M13 rev (–29) sequence (see Table 2). Amplification reactions were performed with an initial denaturation step of 95°C for 4 min and 30 or 35 cycles of 95°C for 2 min, 72°C for 1 min, $38\text{--}48^{\circ}\text{C}$ for 1 min, and a final extension step 72°C for 2 min. If the PCR products were designated for direct sequencing, two subsequent reactions

were performed using the primers prL8-1 and prL8-2, where only one of them contained the M13 rev (–29) sequence (Table 2). The reactions were performed at an annealing temperature of 48 °C. PCR reactions on the partial *lef-9* gene were performed by using primers prL9-1 and prL9-2 with an initial denaturation step of 95 °C for 4 min and 35 cycles of 95 °C for 2 min, 72 °C for 1 min, 45–54 °C for 1 min, and a final extension step 72 °C for 2 min.

Cloning and sequencing of PCR products

The PCR products were directly ligated into the pDrive cloning vector without previous purification using the pDrive cloning kit (Qiagen). For cloning in *Escherichia coli* (strain DH5 α), standard techniques were used. Screening for positive clones was performed by PCR using T7 and SP6 standard primers. Plasmid DNA was isolated from bacteria using the Nucleo Spin Plasmid Kit (Macherey-Nagel GmbH, Düren, Germany). DNA sequencing of both strands was performed by MWG. PCR products used for direct sequencing were purified using the GFX PCR DNA and Gel Band Purification Kit (Amersham, Freiburg, Germany), and both DNA strands were sequenced by MWG or Genterprise (Mainz, Germany).

Sequence alignment and phylogenetic analyses

The deduced amino acid sequences of 30 baculovirus core genes (Herniou et al., 2003; Lange and Jehle, 2003) of 22 fully sequenced baculovirus genomes available from Genbank (Table 1) were individually aligned using ClustalW (Thompson et al., 1994) and subsequently concatenated to a single data set using BioEdit (Hall, 1999). The deduced amino acid sequences of the *polh/gran*, *lef-8*, and *lef-9* gene fragments of the baculoviruses listed in Tables 1 and 3 were aligned and concatenated to a single data set as described above.

Maximum parsimony (MP) phylogenetic trees (1000 bootstrap replicates) were inferred from the amino acid sequence alignments using MEGA, version 2.1 (Kumar et al., 2001) or the PAUP* program, version 4.0b10 (Swofford, 2000). Introduced gaps were treated as missing data; informative characters were treated as multistate, unordered. MP trees were analyzed by a heuristic search with the tree-bisection-reconnection (TBR) branch swapping option. Dayhoff distance corrected neighbor joining (NJ) distance analyses (gamma shape parameter $\alpha = 2.25$) and minimum evolution (ME) analyses (1000 bootstrap replicates, respectively) were performed using MEGA (Nei and Kumar, 2000).

Synonymous and non-synonymous substitution rates based on codon-aligned nucleotide sequences of the partial *lef-8*, *lef-9*, and *polh/gran* genes were calculated according to Nei and Gojobori (1986) using the Synonymous/Non-synonymous Analysis Program (SNAP) at <http://www.hiv.lanl.gov/content/hiv-db/SNAP/>.

Correlation of the AT base composition

The AT nucleotide contents of the whole genomes of 21 lepidopteran-specific NPVs and GVs were correlated with the AT contents of the complete and partial sequences of the *polh/gran*, *lef-8*, and *lef-9* genes and with the weighted arithmetic mean of the three genes. A linear regression was calculated using the least square method. On the basis of the AT contents of the partial *polh/gran*, *lef-8*, and *lef-9* genes, the AT contents of the whole genomes was predicted by using the regression of the weighted arithmetic means.

Acknowledgments

We are indebted to Loy Volkman and the University of California, Berkeley, for providing the historic baculovirus collection of Edward Steinhaus, which was used in part in this study. We also thank Jürg Huber, Federal Biological Research Center, Darmstadt, and Bryony Bonning, Iowa State University, Ames, for sharing baculovirus samples. The authors would be deeply grateful to receive further baculovirus isolates that could be included in an ongoing identification project. We thank Basil Arif and Peter Krell for helpful comments on the manuscript. This research was funded by the Deutsche Forschungsgemeinschaft (grant Je245-4).

References

- Afonso, C.L., Tulman, E.R., Lu, Z., Balinsky, C.A., Moser, B.A., Becnel, J.J., Rock, D.L., Kutish, G.F., 2001. Genome sequence of a baculovirus pathogenic for *Culex nigripalpus*. *J. Virol.* 75, 11157–11165.
- Ahrens, C.H., Russell, R., Funk, C.J., Evans, J.T., Harwood, S.H., Rohrmann, G.F., 1997. The sequence of *Orgyia pseudotsugata* multinucleocapsid nucleopolyhedrosis virus genome. *Virology* 229, 381–399.
- Ayres, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M., Possee, R.D., 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* 202, 586–605.
- Blissard, G., Black, B., Crook, N., Keddle, B.A., Possee, R., Rohrmann, G., Theilmann, D.A., Volkman, L., 2000. Family Baculoviridae. In: van Regenmoertel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B. (Eds.), *Virus Taxonomy—Seventh Report of the International Committee on Taxonomy of Viruses*. Academic Press, San Diego, pp. 195–202.
- Bulach, D.M., Kumar, C.A., Zaia, A., Liang, B., Tribe, D.E., 1999. Group II nucleopolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedrin and DNA polymerase gene sequences. *J. Invertebr. Pathol.* 73, 59–73.
- Chen, X., Ijkel, W.F.J., Tarchini, R., Sun, X., Sandbrink, H., Wang, H., Peters, S., Zuidema, D., Klein-Lankhorst, R., Vlak, J.M., Hu, Z.-H., 2001. The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* 82, 241–257.
- Chen, X., Zhang, W.-J., Wong, G., Chun, A., Lu, B.F., McCutchen, J.K., Presnail, R., Herrmann, M., Dolan, S., Tingey, S., Hu, Z.-H., Vlak, J.M., 2002. Comparative analysis of the complete genome sequence of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. *J. Gen. Virol.* 83, 673–684.

- de Moraes, R.R., Maruniak, J.E., 1997. Detection and identification of multiple baculoviruses using the polymerase chain reaction (PCR) and restriction endonuclease analysis. *J. Virol. Methods* 63, 209–217.
- Federici, B.A., 1997. Baculovirus pathogenesis. In: Miller, L.K. (Ed.), *The Baculoviruses*. Plenum, New York, pp. 33–59.
- Gomi, S., Majima, K., Maeda, S., 1999. Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *J. Gen. Virol.* 80, 1323–1337.
- Gouy, M., Gautier, C., 1982. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 10, 7055–7074.
- Gröner, A., 1986. Specificity and safety of baculoviruses. In: Granados, R.R., Federici, B.A. (Eds.), *The Biology of Baculoviruses*, vol. 1. CRC Press, Boca Raton, pp. 177–201.
- Guarino, L.A., Xu, B., Jin, J.P., Dong, W., 1998. A virus-encoded RNA polymerase purified from baculovirus-infected cells. *J. Virol.* 72, 7985–7991.
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.
- Harrison, R.L., Bonning, B.C., 1999. The nucleopolyhedroviruses of *Rachiplusia ou* and *Anagrapha falcifera* are isolates of the same virus. *J. Gen. Virol.* 80, 2793–2798.
- Harrison, R.L., Bonning, B.C., 2003. Comparative analysis of the genomes of *Rachiplusia ou* and *Autographa californica* multiple nucleopolyhedroviruses. *J. Gen. Virol.* 84, 1827–1842.
- Hashimoto, Y., Hayakawa, T., Ueno, Y., Fujita, T., Sano, Y., Matsumoto, T., 2000. Sequence analysis of the *Plutella xylostella* granulovirus genome. *Virology* 275, 358–372.
- Hayakawa, T., Ko, R., Okano, K., Seong, S., Goto, C., Maeda, S., 1999. Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* 262, 277–297.
- Herniou, E.A., Luque, T., Chen, X., Vlaskin, J.M., Winstanley, D., Cory, J.S., O'Reilly, D.R., 2001. Use of whole genome sequence data to infer baculovirus phylogeny. *J. Virol.* 75, 8117–8126.
- Herniou, E.A., Olezewski, J.A., Cory, J.S., O'Reilly, D.R., 2003. The genome sequence and evolution of baculoviruses. *Annu. Rev. Entomol.* 48, 211–234.
- Herniou, E.A., Olezewski, J.A., O'Reilly, D.R., Cory, J.S., 2004. Ancient coevolution of baculoviruses and their insect hosts. *J. Virol.* 78, 3244–3251.
- Hyink, O., Dellow, R.A., Olsen, M.J., Caradoc-Davies, K.M.B., Drake, K., Herniou, E.A., Cory, J.S., O'Reilly, D.R., Ward, V.K., 2002. Whole genome analysis of the *Epiphyas postvittana* nucleopolyhedrovirus. *J. Gen. Virol.* 83, 957–971.
- Ijkel, W.F.J., van Strien, E.A., Heldens, J.G.M., Broer, R., Zuidema, D., Goldbach, R.W., Vlaskin, J.M., 1999. Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* 80, 3289–3304.
- Jehle, J.A., 2004. The mosaic structure of the polyhedrin gene of the *Autographa californica* nucleopolyhedrovirus (AcMNPV). *Virus Genes* 29, 5–8.
- Kumar, S., Koichiro, T., Jacobsen, I.B., Nei, M., 2001. MEGA2: Molecular Evolutionary Genetics Analysis Software. Arizona State University, Tempe, AZ, USA.
- Kuzio, J., Pearson, M.N., Harwood, S.H., Funk, C.J., Evans, J.T., Slavicek, J.M., Rohrmann, G.F., 1999. Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. *Virology* 253, 17–34.
- Lange, M., Jehle, J.A., 2003. The genome of the *Cryptophlebia leucotreta* granulovirus. *Virology* 317, 220–236.
- Li, Q., Donly, C., Li, L., Willis, L.G., Theilmann, D.A., Erlandson, M.A., 2002a. Sequence and organization of the *Mamestra configurata* nucleopolyhedrovirus genome. *Virology* 294, 106–121.
- Li, L., Donly, C., Li, Q., Willis, L.G., Keddie, B.A., Erlandson, M.A., Theilmann, D.A., 2002b. Identification and genomic analysis of a second species of nucleopolyhedrovirus isolated from *Mamestra configurata*. *Virology* 297, 226–244.
- Luque, L., Finch, R., Crook, N., O'Reilly, D.R., Winstanley, D., 2001. The complete sequence of the *Cydia pomonella* granulovirus genome. *J. Gen. Virol.* 82, 2531–2547.
- Martignoni, M.E., Iwai, P.J., 1981. A catalogue of viral diseases of insects, mites and ticks. In: Burges, H.D. (Ed.), *Microbial Control of Pests and Plant Diseases 1970–1980*. Academic Press, London, pp. 897–911.
- Moser, B.A., Becnel, J.J., White, S.E., Alfonso, C., Kutish, G., Shanker, S., Almira, E., 2001. Morphological and molecular evidence that *Culex nigripalpus* baculovirus is an unusual member of the family Baculoviridae. *J. Gen. Virol.* 82, 283–297.
- Nakai, M., Goto, C., Kang, W.-K., Shikata, M., Luque, T., Kunimi, Y., 2003. Genome sequence and organization of a nucleopolyhedrovirus isolated from the smaller tea tortrix. *Adoxophyes honmai*. *Virology* 316, 171–183.
- Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426.
- Nei, M., Kumar, S., 2000. *Molecular Evolution and Phylogenetics*. Oxford Univ. Press, New York.
- 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.
- Pearson, M.N., Rohrmann, G.F., 2002. Transfer, incorporation, and substitution of envelope fusion proteins among members of the Baculoviridae, Orthomyxoviridae, and Metaviridae (insect retrovirus) families. *J. Virol.* 304, 81–88.
- Rohrmann, G.F., 1986. Polyhedrin structure. *J. Gen. Virol.* 67, 1499–1513.
- Steinhaus, E.A., Marsh, G.A., 1962. Report of diagnoses of diseased insects 1951–1961. *Hilgardia* 33, 349–490.
- Swofford, D.L., 2000. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), Edition 4.0. Sinauer Associates, Sunderland.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680.
- van Regenmoertel, M.H.V., 2000. Introduction to the species concept in virus taxonomy. In: van Regenmoertel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B. (Eds.), *Virus Taxonomy—Seventh Report of the International Committee on Taxonomy of Viruses*. Academic Press, San Diego, pp. 3–16.
- Wormleaton, S., Kuzio, J., Winstanley, D., 2003. The complete sequence of the *Adoxophyes orana* granulovirus genome. *Virology* 311, 350–365.
- Zanotto, P.M., Kessing, B.D., Maruniak, J.E., 1993. Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. *J. Invertebr. Pathol.* 62, 147–164.