Journal of General Virology (2003), **84**, 1827–1842 DOI 10.1099/vir.0.19146-0

Comparative analysis of the genomes of *Rachiplusia ou* and *Autographa californica* multiple nucleopolyhedroviruses

Robert L. Harrison and Bryony C. Bonning

Department of Entomology and Interdepartmental Program in Genetics, Iowa State University, Ames, Iowa 50011, USA

Correspondence
Bryony C. Bonning
bbonning@iastate.edu

Received 5 February 2003
Accepted 20 March 2003

The *Rachiplusia ou* multiple nucleopolyhedrovirus (RoMNPV) is a variant of *Autographa californica* MNPV (AcMNPV) but is significantly more virulent against several major agricultural pests. The genome sequence of the R1 strain of RoMNPV was determined and compared to that of AcMNPV strain C6. The RoMNPV genome is approximately 131·5 kbp with a G+C content of 39·1 %. The homologous repeat regions (*hr*s) described for AcMNPV-C6 are present in RoMNPV-R1 but the *hr*s of RoMNPV have fewer palindromic repeats. The RoMNPV-R1 nucleotide sequence is almost completely collinear with the sequence of AcMNPV-C6 and contains homologues of 150 of the 155 ORFs described for AcMNPV-C6. Deletions, insertions and substitutions have resulted in the loss of homologues for AcMNPV ORFs *ac2* (*bro*), *ac3* (*ctl*), *ac97*, *ac121* and *ac140* from the RoMNPV genome. The average amino acid sequence identity between RoMNPV and AcMNPV ORFs is 96·1 % and there are differences in promoter motif composition for 23 of these ORFs. Maximum-likelihood analysis of selection pressures on AcMNPV and RoMNPV ORFs indicate that ORFs *ro18*/*ac20-ac21* (*arif-1*) and *ro135*/*ac143* (*odv-e18*) have undergone positive selection.

INTRODUCTION

Baculoviruses are invertebrate-specific viruses with large, double-stranded DNA genomes contained within enveloped, rod-shaped virions. The baculoviruses are grouped into one family, the *Baculoviridae*, with two genera, *Nucleopolyhedrovirus* and *Granulovirus*, which are distinguished by the morphology of the virus occlusion bodies. Members of the *Baculoviridae* have been isolated from insects mainly within the order Lepidoptera (Adams & McClintock, 1991; Blissard *et al.*, 2000). Because nucleopolyhedroviruses (NPVs) are effective against many lepidopteran pests and do not infect non-target organisms, they have been the subject of intensive study as environmentally benign insecticides. Wild-type baculoviruses have been effective when deployed against agricultural and forestry pests (Cunningham, 1995; Moscardi, 1999). Genetic engineering of NPVs to shorten the survival time of infected hosts has improved the performance of NPV-based insecticides in the field (Treacy & All, 1996; Smith *et al.*, 2000; Treacy *et al.*, 2000). Further studies on the molecular mechanisms of baculovirus infection, replication and host specificity are essential for additional augmentation of insecticidal potential.

Rachiplusia ou multiple nucleopolyhedrovirus (RoMNPV) was first identified in 1960 when it caused an epizootic in populations of the mint looper, *Rachiplusia ou*, in mint fields (Paschke & Hamm, 1961). Restriction mapping and nucleic acid hybridization studies revealed that RoMNPV is closely related to *Autographa californica* MNPV (AcMNPV), the type species for *Nucleopolyhedrovirus* (Jewell & Miller, 1980; Smith & Summers, 1980, 1982). The sequence of the *Eco*RI-G restriction fragment of RoMNPV (R1 strain) confirmed that a high degree of nucleotide sequence identity exists between RoMNPV and AcMNPV in this region (Harrison & Bonning, 1999). In addition, the RoMNPV *Eco*RI-G sequence was found to be almost completely identical to the sequence of the corresponding region in *Anagrapha falcifera* MNPV (AfMNPV; Federici & Hice, 1997), which was first identified in 1985 (Hostetter & Puttler, 1991). Restriction enzyme analysis and bioassays against three different host species demonstrated that AfMNPV and RoMNPV are isolates of the same virus (Harrison & Bonning, 1999). RoMNPV and AfMNPV are characterized as variants of AcMNPV (Blissard *et al.*, 2000).

Both AcMNPV and RoMNPV/AfMNPV are known to infect a relatively large number of lepidopteran species (31 and 43 species, respectively; Granados & Williams, 1986; Payne, 1986; Hostetter & Puttler, 1991). Although the host ranges for these viruses overlap, there are significant differences in the abilities of AcMNPV and RoMNPV/AfMNPV to infect and kill several agricultural pest species. The corn earworm,

The GenBank accession number of the sequence reported in this paper is AY145471.

Helicoverpa zea, is approximately 2·5- to 29-fold more susceptible to RoMNPV/AfMNPV than AcMNPV (Harrison & Bonning, 1999; Hostetter & Puttler, 1991). *Ostrinia nubilalis*, the European corn borer, is approximately 5- to 11-fold more susceptible to RoMNPV than AcMNPV (Harrison & Bonning, 1999; Lewis & Johnson, 1982). The navel orangeworm, *Amyelois transitella*, is non-permissive to AcMNPV but can be infected and killed with AfMNPV (Vail *et al.*, 1993; Cardenas *et al.*, 1997). The tobacco hornworm, *Manduca sexta*, a species that is highly refractory to AcMNPV (Washburn *et al.*, 2000), is susceptible to a dose of 100 polyhedra mm^{-2} diet surface of AfMNPV (Hostetter & Puttler, 1991). The fall armyworm, *Spodoptera frugipera*, and the velvet bean caterpillar *Anticarsia gemmatalis*, are also more susceptible to AfMNPV than AcMNPV (Hostetter & Puttler, 1991).

Given the high degree of sequence similarity between the genomes of these viruses, we reasoned that comparison of the genomes of RoMNPV and AcMNPV would reveal genetic differences that could account for differences in host range and virulence. Towards this end, we sequenced the RoMNPV genome. Here we present a comparison of the RoMNPV and AcMNPV genomes and an analysis of selection pressures on RoMNPV and AcMNPV genes.

METHODS

Virus and cells. RoMNPV-R1 (Smith & Summers, 1980) was propagated in *S. frugiperda* (Sf) cell lines (Vaughn *et al.*, 1977) and titrated by plaque assay. Sf21 cells were grown in Ex-Cell 405 medium (JRH Biosciences) supplemented with 3 % FBS (Intergen) and antibiotics (1 U penicillin ml^{-1} and 1 μg streptomycin ml^{-1}; Sigma). Sf9 cells were grown in TNM-FH medium (JRH Biosciences) supplemented with 3 % FBS, antibiotics and 0·1 % Pluronic F-68 (JRH Biosciences).

Viral DNA isolation and cloning. Sf9 or Sf21 cells were infected with RoMNPV-R1 at an m.o.i. of 1. Budded virus (BV) was harvested at 5 days post-infection (p.i.). BV was precipitated by overnight incubation on ice with an equal volume of 20 % polyethylene glycol and 1 M NaCl. After pelleting by centrifugation, the BV was resuspended in 10 mM Tris/HCl and 1 mM EDTA (pH 8·0) and incubated for 3 h at 37 °C with 1 % SDS and 1 mg proteinase K ml^{-1}. Viral DNA was purified by phenol/chloroform extraction and ethanol precipitation. Alternatively, the BV pellet was resuspended in 1 ml DNAzol, a genomic DNA isolation reagent (Invitrogen/Life Technologies), and viral DNA was isolated following the manufacturer's instructions.

RoMNPV-R1 DNA was digested with *Eco*RI, *Hin*dIII, *Pst*I, *Bgl*II and *Xba*I. Restriction fragments were ligated into the vectors pGEM-9Zf(−) (Promega), pUC18 and pUC19M (Clontech), a variant of pUC19 in which the *Eco*RI site has been replaced with an *Eco*RV site. Ligation products were transformed into competent *Escherichia coli* JM109. Plasmid DNA for clones with RoMNPV inserts was prepared using Qiagen columns (Qiagen).

DNA sequencing and sequence analysis. A 'primer walking' strategy was used to sequence selected RoMNPV restriction fragments from both ends. Automated dideoxy sequencing was performed at the Iowa State University DNA Sequencing and Synthesis Facility. Reactions were set up using the Applied Biosystems Prism BigDye Terminator Cycle Sequencing kit with Ampli*Taq* DNA polymerase and electrophoresed on an Applied Biosystems Prism 377 DNA sequencer.

To confirm the order of some RoMNPV-R1 restriction fragments, regions encompassing restriction fragment junctions were amplified from viral DNA by PCR, purified with Qiagen columns and sequenced. The sequences of selected AcMNPV strain C6 ORFs were also re-determined by amplifying the ORFs from an AcMNPV-C6 stock and sequencing the amplification products.

DNA sequence data were compiled and analysed with the software of the Wisconsin package (version 10.0, Genetics Computer Group) and the Lasergene suite (DNASTAR). ORFs greater than 50 codons in length that did not overlap larger ORFs by more than 75 nt were selected for further characterization. Predicted amino acid sequence identities were obtained from the results of protein database searches using the standard protein–protein BLAST algorithm (http://www.ncbi.nlm.nih.gov/blast/).

To assess the relationship of the AcMNPV and RoMNPV polyhedrins to the polyhedrins of other baculoviruses, phylogenetic analysis of 35 baculovirus occlusion matrix proteins was carried out. Amino acid sequences in this data set were aligned with CLUSTAL W (Thompson *et al.*, 1994) using Gonnet matrices with a gap penalty of 15 and a gap extension penalty of 0·3 and adjusted manually. Phylogenetic inferences were performed with MEGA, version 2.1 (Kumar *et al.*, 2001) using minimum-evolution (ME) and maximum-parsimony (MP) methods (Nei & Kumar, 2000). ME and MP trees were sought using a close-neighbour-interchange heuristic search, starting with one initial tree generated by the neighbour-joining method (for ME) or 10 initial trees generated by random addition of sequences (for MP). For ME, evolutionary distance was estimated using the gamma distance model with the gamma shape parameter set to 2·25. The reliability of the trees was tested with the bootstrap resampling strategy using 1000 replicates. For comparison, ME trees of NPV *dnapol* and *p10* predicted amino acid sequences were constructed in the same way using alignments assembled with a gap penalty of 10 and a gap extension penalty of 0·2.

Analysis of selection pressures on individual genes. PAML software (Yang, 1997; http://abacus.gene.ucl.ac.uk/software/paml.html) was used to investigate selection pressures on the genes of AcMNPV and RoMNPV. This software uses a maximum-likelihood approach to determine the numbers of non-synonymous (amino acid changing) substitutions per non-synonymous site (d_N) and of synonymous (silent) substitutions per synonymous site (d_S). The ratio of d_N to d_S, ω, is a measure of the selective pressure on a gene. Genes with $\omega = 1$ are undergoing neutral evolution, in which there is no effect of non-synonymous mutations on fitness. Genes with $\omega < 1$ are undergoing negative or purifying selection, in which non-synonymous mutations are deleterious and are eliminated at a faster rate than synonymous mutations. Genes with $\omega > 1$ are undergoing positive or diversifying selection, in which non-synonymous mutations are favourable and are fixed at a faster rate than synonymous mutations.

All ORFs common to AcMNPV and RoMNPV were analysed initially using a pairwise comparison method that assumes a single value of ω for all codon sites in an ORF. A subset of these ORFs that are present in multiple genomes or known to be expressed was analysed further using models that allow for heterogeneous values of ω among codon sites (Yang *et al.*, 2000). With this second analysis, ORF alignments were fitted to six models: (1) M0 assumes one ω value for all codons; (2) M1 divides codons into an invariant category p_0, where ω is set at zero (purifying selection), and a neutral category p_1, where ω is set at one (neutral evolution); (3) M2 includes p_0 and p_1 from M1 and adds a third category p_2, where ω is estimated from the underlying data and

can be greater than one; (4) M3 divides codons among three categories of sites (p_0, p_1 and p_2). ω is estimated independently for all three categories and can be greater than one; (5) M7 features 10 categories modelled with a discrete β distribution. The shape of the distribution is determined by parameters p and q, and ω values for these categories cannot be greater than one; and (6) M8 includes the 10 categories of M7 (collectively referred to as p_0), and uses an additional category p_1, where ω can be greater than one.

Models M0 and M1 are nested with models M2 and M3, and model M7 is nested with M8. Models that are nested together can be compared statistically using a likelihood ratio test, in which twice the difference between the log-likelihood values for two models is compared with a χ^2 distribution table with the degrees of freedom equal to the difference in the number of parameters between the two models. This comparison supplies a P value for the probability that the null hypothesis (no positive selection, embodied in models M0, M1 and M7) is an equally good or better fit for the data when compared to the nested models that indicate positive selection. Positive selection can be inferred from this analysis when: (1) models M2, M3 or M7 indicate a group of codons with an ω ratio greater than 1; and (2) the likelihood of the positive selection model is significantly higher than that of the nested null hypothesis model (at $P<0{\cdot}05$). An empirical Bayes procedure calculates the probabilities for individual codons belonging to each of the site categories and can be used to infer which codons are under positive selection.

For selection pressure analysis, predicted amino acid sequences of AcMNPV and RoMNPV ORFs were aligned using CLUSTAL W, as described for phylogenetic analysis of occlusion matrix proteins. The sequences in the alignment were then converted back to the original nucleotide sequences. The CODEML and CODEMLSITES programs of PAML were run with the nucleotide sequence alignments. For pairwise analysis, codon frequency bias was accounted for using the F61 model of codon frequency, in which the frequency of each codon is used as a free parameter. Analysis with models allowing ω to vary used the $F3\times4$ model, in which codon frequencies are calculated from average nucleotide frequencies at the three codon positions. In all analyses, the transition/transversion ratio (κ) was estimated from the underlying data.

Alignments and output files from these analyses can be downloaded from http://www.ent.iastate.edu/dept/faculty/bonningb/selection_pressure.zip.

RESULTS AND DISCUSSION

General characteristics of the RoMNPV genome

The size of the RoMNPV-R1 genome is 131 526 bp, 2368 bp smaller than the AcMNPV-C6 genome (Fig. 1). This size difference is accounted for almost entirely by the absence of the *ac2* (*bro*)-*ac3* (*ctl*) region from RoMNPV (Harrison & Bonning, 1999) and the reduced size of RoMNPV homologous repeat regions (*hr*s). The RoMNPV G+C content (39·1 %) is slightly less than that of AcMNPV (41 %) and *Bombyx mori* NPV (BmNPV; 40 %). The RoMNPV genome is almost completely collinear with the AcMNPV-C6 sequence and the overall nucleotide sequence identity between RoMNPV and AcMNPV is approximately 96 %. For ease of comparison with the AcMNPV-C6 genome, nucleotide position #1 of RoMNPV was set to the nucleotide that aligned with the first nucleotide of the AcMNPV-C6 sequence. RoMNPV has no extra sequences that are not also present in the AcMNPV genome. Table 1 lists RoMNPV homologues of 150 expressed or potentially expressed AcMNPV-C6 ORFs described by Ayres *et al.* (1994). These homologues are in the same relative positions and orientations as the AcMNPV ORFs. In addition, Table 1 lists three RoMNPV homologues of ORFs previously undocumented in AcMNPV (*ac15a*, *ac100a* and *ac139a*). These ORFs overlap larger, adjacent ORFs by less than 75 bp. They are not preceded by baculovirus early or late gene promoter motifs in either virus and do not contain any recognizable conserved domains in their predicted amino acid sequences. With the exception of *ro13a*/*ac15a* (which has 55·4 % sequence identity with BmNPV small ORF7a; Gomi *et al.*, 1999), these ORFs share no significant sequence identity with other viral or cellular genes.

RoMNPV *hr*s

*hr*s are sequences that function as enhancers of gene expression and origins of DNA replication (Possee & Rohrmann, 1997). RoMNPV has nine *hr*s consisting of imperfect 30 bp palindromic repeats similar to those described for AcMNPV *hr*s (Fig. 2). The RoMNPV *hr*s occupy the same positions on the genome with respect to surrounding ORFs as the AcMNPV *hr*s (Fig. 1) and are numbered in the same order as the AcMNPV *hr*s. Each of the RoMNPV *hr*s has at least one less palindromic repeat than the corresponding AcMNPV *hr*s (Fig. 3). RoMNPV *hr2* displays the greatest reduction in size, with four less repeats than AcMNPV *hr2*. The spacing between the palindromic repeats of RoMNPV and AcMNPV *hr*s is variable (Fig. 3) and the alignments of AcMNPV and RoMNPV *hr*s require multiple gaps, suggesting that numerous deletions and insertions have occurred in the regions between the palindromic repeats in the *hr*s of these viruses. A similar pattern of deletions and insertions has been observed upon comparison of the *hr*s of *Helicoverpa armigera* single nucleopolyhedrovirus (HaSNPV) and *Helicoverpa zea* SNPV (HzSNPV; Chen *et al.*, 2002), and *Mamestra configurata* NPV (McNPV) isolates 90/2 and 96B (Li *et al.*, 2002a). These observations confirm previous suggestions that *hr*s are sites of frequent recombination and rearrangement in baculovirus genomes.

Comparison of RoMNPV and AcMNPV ORFs

There were five fewer potentially expressed ORFs identified in RoMNPV than in AcMNPV. This reduced ORF count is due to the truncation of three ORFs and the complete deletion of two ORFs.

As reported previously (Federici & Hice, 1997; Harrison & Bonning, 1999), RoMNPV is missing a 1275 bp region, which, in AcMNPV, contains *ac2* (baculovirus repeated ORF, *bro*) and *ac3* (conotoxin-like gene, *ctl*). No other homologues of these genes were detected elsewhere in the RoMNPV genome. The *ctl* ORF has also been found in the genomes of *Orgyia pseudotsugata* MNPV (OpMNPV; Ahrens *et al.*, 1997), McNPV-90/2 and -96B (Li *et al.*, 2002a, b), *Lymantria dispar* MNPV (LdMNPV; Kuzio *et al.*,

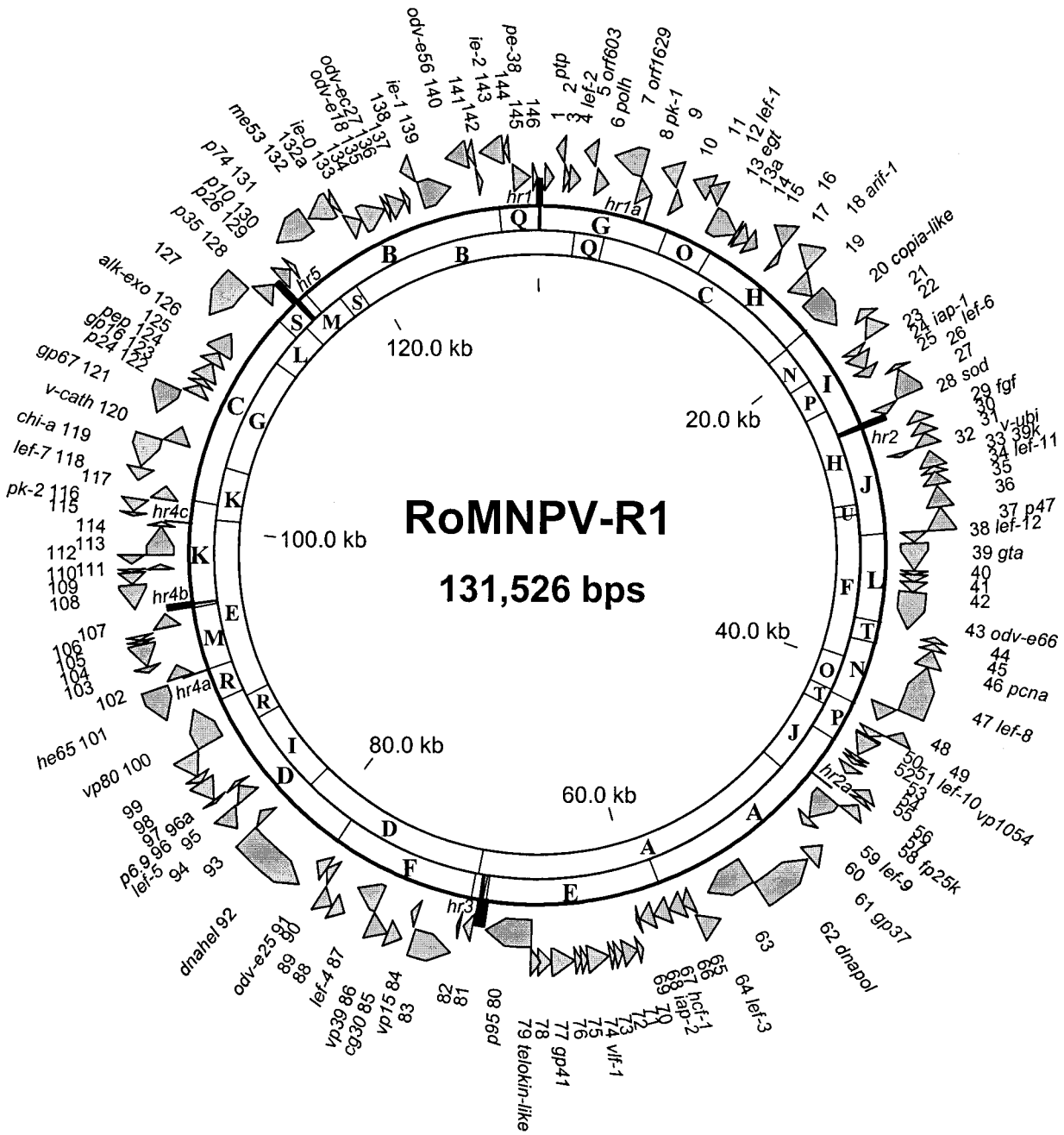

Fig. 1. Circular map of the RoMNPV genome (R1 strain). Locations for *Hin*dIII and *Eco*RI are shown on the inner and outer rings, respectively. The positions for the 149 ORFs listed in Table 1 are presented as arrowheads, with the direction of the arrowhead indicating the orientation of the ORF. The locations for the nine homologous repeat regions (*hr*s) are indicated.

1999) and *Xestia c-nigrum* granulovirus (XecnGV; Hayakawa *et al.*, 1999). Deletion of *ctl* from AcMNPV had no effect on replication in tissue culture (Eldridge *et al.*, 1992) and the function of *ctl* is unknown. Homologues of the *bro* ORF are present in multiple copies in other baculovirus genomes. The BRO proteins of BmNPV are associated with nucleoprotein complexes and bind nucleic acids (Zemskov *et al.*, 2000). Kang *et al.* (1999) were unable to produce a viable BmNPV *bro-d* single deletion mutant or a *bro-a*/*bro-c* double deletion mutant, suggesting that these ORFs may be essential for BmNPV replication in cell culture (Kang *et al.*, 1999). *bro* genes are also absent from *Spodoptera exigua* NPV (SeNPV; IJkel *et al.*, 1999) and *Plutella xylostella* GV (PxGV; Hashimoto *et al.*, 2000).

Three small ($\leqslant$60 codons) AcMNPV ORFs are also missing

Table 1. Potentially expressed ORFs of RoMNPV (R1 strain)

Of the 149 ORFs listed, 146 are homologues of AcMNPV ORFs described previously. Four RoMNPV ORFs are homologues of AcMNPV ORFs that are split in two in the original C6 sequence. These RoMNPV ORFs are, therefore, homologous to two AcMNPV ORFs each in the original C6 sequence. Hence, for the 155 ORFs described for C6, there are 150 (146 + 4) homologues present in RoMNPV.

					Comparison with AcMNPV ORFs			
ORF	Name	Position*	aa (Da)†	Promoter motifs‡	AcMNPV homologue (aa)	% identity (range of alignment)	Promoter motifs‡	ω (d_N/d_S)§
1	*ptp*	392→898	168 (19 330)	E, L	*ac1* (168)	98·8 (166/168)	E, L	0·031
2		962←1417	151 (17 556)	E	*ac4* (151)	96 (145/151)	E	0·072
3		1446→1775	109 (12 467)	L	*ac5* (109)	98·2 (107/109)	L	0·103
4	*lef-2*	1756→2388	210 (23 741)	E, L	*ac6* (210)	95·7 (201/210)	E, L	0·245
5	*orf603*	2429←3064	211 (24 504)	L	*ac7* (201)	87·1 (183/210)		0·239
6	*polh*	3230→3967	245 (28 793)	L	*ac8* (245)	90·2 (221/245)	L	0·071
7	*orf1629capsid*	3997←5625	542 (60 677)		*ac9* (543)	93·2 (510/547)		0·205
8	*pk-1*	5624→6442	272 (32 074)	L	*ac10* (272)	97·4 (265/272)	L	0·031
	hr1a	6454–6491						
9		6515←7534	339 (39 926)	E	*ac11* (340)	95·3 (322/338)	E, C	0·196
10		7649→8227	192 (22 710)	E	*ac12* (217)	93·1 (175/188)	E	0·394
11		8253←9239	328 (38 760)	L	*ac13* (327)	95·4 (313/328)	L	0·150
12	*lef-1*	9128←9931	267 (30 975)		*ac14* (266)	95·1 (254/267)		0·239
13	*egt*	10045→11565	506 (57 279)		*ac15* (506)	96·6 (489/506)		0·136
13a		11578→11745	55 (5 975)		*ac15a* (55)	94·5 (52/55)		NC
14		11711→12388	225 (25 898)		*ac16* (225)	97·8 (220/225)		0·095
15		12357→12989	210 (24 101)		*ac17* (209)	96·7 (203/210)		0·342
16		13019←14077	352 (40 819)	E	*ac18* (353)	98·0 (346/353)	E	0·071
17		14079→14405	108 (12 161)	L	*ac19* (108)	98·1 (106/108)	L	0·077
18	*arif-1*	14608←15849	413 (47 413)	C, L	*ac20-ac21* (417)	94·7 (395/417)	C	0·189
19		15886→17034	382 (43 681)		*ac22* (382)	98·4 (376/382)	L	0·079
20	*copia-like env*	17098→19161	687 (79 740)	L	*ac23* (690)	95·2 (657/690)	E, L	0·204
21		19210←19716	168 (19 189)		*ac24* (169)	97·0 (162/167)	L	0·143
22		19757←20704	315 (36 501)		*ac25* (316)	98·7 (312/316)		0·036
23		20780→21169	129 (14 689)	E, L	*ac26* (129)	96·1 (124/129)	E, L	0·283
24	*iap-1*	21171→22031	286 (33 241)	E	*ac27* (286)	95·5 (273/286)	E	0·167
25	*lef-6*	22036→22560	174 (20 616)	E	*ac28* (173)	93·0 (160/172)	E	0·186
26		22611←22826	71 (8 521)		*ac29* (71)	95·8 (68/71)		0·315
27		22880←24271	463 (54 715)	E	*ac30* (463)	95·5 (442/463)		0·265
28	*sod*	24332→24787	151 (16 256)	L	*ac31* (151)	97·4 (147/151)	L	0·121
	hr2	24805–25141						
29	*fgf*	25216←25761	181 (20 694)		*ac32* (181)	97·2 (176/181)		0·108
30		25880←26428	182 (20 980)	E, L	*ac33* (182)	94·5 (171/181)	E, L	0·213
31		26441←27088	215 (24 874)	L	*ac34* (215)	99·5 (214/215)	L	0·056
32	*v-ubi*	27109→27342	77 (8 652)	L	*ac35* (77)	100·0 (77/77)	L	NC
33	*39K*	27387←28214	275 (31 498)	C, L	*ac36* (275)	93·1 (256/275)	L	0·252
34	*lef-11*	28208←28546	112 (13 260)	E	*ac37* (112)	95·5 (107/112)	E	0·505
35		28509←29159	216 (25 300)	L	*ac38* (216)	97·2 (210/216)	L	0·114
36		29239←30327	362 (43 553)		*ac39* (363)	93·0 (333/358)		0·254
37	*p47*	30335←31540	401 (47 490)	E	*ac40* (401)	97·5 (391/401)	E	0·190
38	*lef-12*	31539→32084	181 (21 198)	E	*ac41* (181)	95·6 (173/181)	E	0·107
39	*gta*	32159→33676	505 (58 986)	E	*ac42* (506)	97·2 (492/506)		0·117
40		33689→33922	77 (8 847)	L	*ac43* (77)	96·1 (74/77)	L	0·156
41		33903→34298	131 (14 986)	E	*ac44* (131)	95·4 (125/131)	E	0·182
42		34300→34875	191 (22 684)		*ac45* (192)	92·7 (178/192)		0·257
43	*odv-e66*	34860→36965	701 (78 857)	L	*ac46* (704)	96·9 (682/704)	L	0·134
44		37083←37352	89 (10 561)	C	*ac47* (88)	93·3 (83/89)	C	0·181

Table 1. (*cont.*)

ORF	Name	Position*	aa (Da)†	Promoter motifs‡	Comparison with AcMNPV ORFs			
					AcMNPV homologue (aa)	% identity (range of alignment)	Promoter motifs‡	ω (d_N/d_S)§
45		37426←37767	113 (12 964)	E	*ac48* (113)	93·8 (106/113)	E	0·255
46	*pcna*	37791←38648	285 (32 081)		*ac49* (285)	94·7 (270/285)		0·239
47	*lef-8*	38671←41301	876 (101 748)		*ac50* (876)	97·9 (858/876)		0·070
48		41328→42284	318 (37 617)		*ac51* (318)	95·3 (303/318)		0·278
49		42275←42859	194 (23 308)		*ac52* (194)	96·4 (187/194)	E	0·159
50		42861→43280	139 (17 007)	L	*ac53* (139)	97·8 (136/139)	L	0·071
51	*lef-10*	43277→43513	78 (8 568)	L	*ac53a* (78)	96·2 (75/78)	L	0·179
52	*vp1054*	43371→44468	365 (42 193)	L	*ac54* (365)	98·4 (359/365)		0·130
53		44560→44781	73 (8 136)		*ac55* (73)	97·3 (71/73)		0·237
54		44783→45037	84 (9 857)	L	*ac56* (84)	100·0 (84/84)	L	NC
55		45222→45707	161 (18 951)	E, L	*ac57* (161)	98·1 (158/161)	E	0·051
56		45725←46243	172 (20 326)		*ac58-ac59* (172)	97·7 (168/172)		0·128
57		46255←46518	87 (10 147)	L	*ac60* (87)	100·0 (87/87)	L	NC
58	*fp25k*	46660←47304	214 (25 183)	L	*ac61* (214)	99·1 (212/214)	L	0·038
	hr2a	46826–46855						
59	*lef-9*	47323→48882	519 (59 728)		*ac62* (516)	99·2 (512/516)		0·176
60		48943→49413	156 (18 558)	C	*ac63* (155)	97·4 (150/154)	C	0·190
61	*gp37*	49430←50338	302 (34 870)	L	*ac64* (302)	96·4 (291/302)	L	0·181
62	*dnapol*	50482←53439	985 (114 558)	E	*ac65* (984)	96·8 (953/985)	E	0·099
63		53448→55880	810 (94 066)	L	*ac66* (808)	94·8 (769/811)	L	0·206
64	*lef-3*	55883←57040	385 (44 574)		*ac67* (385)	97·7 (376/385)		0·092
65		56882→57460	192 (22 444)		*ac68* (192)	95·3 (183/192)		1·20
66		57438→58226	262 (30 300)		*ac69* (262)	96·6 (253/262)		0·113
67	*hcf-1*	58272→59144	290 (34 467)	C	*ac70* (290)	84·1 (244/290)		0·600
68	*iap-2*	59178→59927	249 (28 708)	L	*ac71* (249)	94·8 (236/249)	L	0·172
69		59986→60168	60 (7 081)	E, L	*ac72* (60)	98·3 (59/60)	E, L	NC
70		60177←60476	99 (11 579)	E, L	*ac73* (99)	98·0 (97/99)	E, L	0·056
71		60473←61261	262 (30 447)	L	*ac74* (265)	95·1 (252/265)	L	0·190
72		61279←61680	133 (15 511)	L	*ac75* (133)	100·0 (133/133)	L	NC
73		61696←61950	84 (9 439)	E, L	*ac76* (84)	100·0 (84/84)	E, L	NC
74	*vlf-1*	61966←63108	380 (44 478)	L	*ac77* (379)	99·2 (377/380)	L	0·019
75		63115←63444	109 (12 556)	L	*ac78* (109)	99·1 (108/109)	L	0·030
76		63447←63761	104 (12 198)		*ac79* (104)	95·2 (99/104)		0·478
77	*gp41*	63764←64993	409 (45 450)	L	*ac80* (409)	99·0 (405/409)	L	0·053
78		64983←65684	233 (26 926)	E, L	*ac81* (233)	97·0 (226/233)	E, L	0·210
79	*telokin-like*; BV-associated	65533←66075	180 (19 759)		*ac82* (180)	95·0 (171/180)	L	0·856
80	*p95capsid*	66041→68611	856 (97 520)	L	*ac83* (847)	95·4 (814/853)	L	0·190
	hr3	68628–69279						
81		69365→69859	164 (19 028)	C	*ac84* (188)	89·4 (144/161)		0·436
82		70062→70223	53 (6 366)		*ac85* (53)	100·0 (53/53)		NC
83		70259←72343	694 (80 775)	C	*ac86* (694)	99·9 (693/694)	C	0·135
84	*vp15*	72484→72864	126 (15 047)	E	*ac87* (126)	100·0 (126/126)	E	NC
85	*cg30*	72865←73659	264 (30 092)		*ac88* (264)	100·0 (264/264)		NC
86	*vp39capsid*	73662←74705	347 (39 008)	L	*ac89* (347)	98·3 (341/347)	L	0·081
87	*lef-4*	74724→76118	464 (54 060)		*ac90* (464)	97·2 (451/464)		0·119
88		76115←76828	237 (25 416)	L	*ac91* (224)	92·0 (219/238)	L	NC
89		76866←77645	259 (30 936)		*ac92* (259)	100·0 (259/259)		NC
90		77644→78129	161 (18 374)		*ac93* (161)	98·1 (158/161)		0·099
91	*odv-e25*	78137→78823	228 (25 491)	E, L	*ac94* (228)	97·8 (223/228)	E, L	0·044
92	*dnahel*	78859←82524	1221 (143 233)	E, L	*ac95* (1221)	98·2 (1199/1221)	E, L	0·048
93		82511→83032	173 (19 853)		*ac96* (173)	97·7 (169/173)		0·228

Table 1. (*cont.*)

ORF	Name	Position*	aa (Da)†	Promoter motifs‡	Comparison with AcMNPV ORFs			
					AcMNPV homologue (aa)	% identity (range of alignment)	Promoter motifs‡	ω (d_N/d_S)§
94		83158←84120	320 (37 996)	E, L	*ac98* (320)	98·8 (316/320)	E, L	0·082
95	*lef-5*	84055→84852	265 (31 107)		*ac99* (265)	98·1 (260/265)		0·072
96	*p6·9*	84849←85016	55 (6 884)	L	*ac100* (55)	100·0 (55/55)	L	NC
96a		84887→85111	74 (8 395)		*ac100a* (74)	100 (74/74)		NC
97		85058←86143	361 (41 553)	L	*ac101* (361)	99·2 (358/361)	L	0·039
98		86162←86527	121 (13 229)	L	*ac102* (122)	95·1 (116/122)	E, L	0·188
99		86508←87671	387 (45 392)	L	*ac103* (387)	98·4 (381/387)	L	0·055
100	*vp80capsid*	87698→89779	693 (80 183)	L	*ac104* (691)	97·0 (672/693)	L	0·156
101	*he65*	89807←91522	572 (67 607)		*ac105* (553)	95·1 (524/551)	C, L	0·209
	hr4a	91594–91623						
102		91691→92428	245 (28 420)	E, L	*ac106-ac107* (243)	96·3 (236/245)	L	0·040
103		92429←92746	105 (11 873)	L	*ac108* (105)	97·1 (101/104)	L	0·134
104		92758←93930	390 (44 865)	E, L	*ac109* (390)	99·0 (386/390)	E, L	0·030
105		93966←94136	56 (6 817)	L	*ac110* (56)	98·2 (55/56)	L	0·068
106		94185←94388	67 (8 221)		*ac111* (67)	88·1 (59/67)	E	0·707
107		94554→95318	254 (30 474)	C	*ac112-ac113* (258)	94·2 (243/258)	E, C	0·258
	hr4b	95394–95750						
108		95755←97032	425 (49 353)	E, L	*ac114* (424)	97·2 (413/425)	E, L	0·080
109		97054←97668	204 (23 097)	E, L	*ac115* (204)	96·6 (197/204)	E, L	0·190
110		97676←97846	56 (6 434)		*ac116* (56)	87·5 (49/56)		3·50
111		97782→98063	93 (10 753)	C	*ac117* (95)	95·7 (88/92)	C	0·354
112		98097←98570	157 (19 008)	E	*ac118* (157)	93·6 (147/157)		0·350
113		98565→100157	530 (59 772)	L	*ac119* (530)	97·2 (515/530)	L	0·081
114		100162→100410	82 (9 481)	L	*ac120* (82)	92·7 (76/82)	L	0·288
	hr4c	100470–100499						
115		100574←100762	62 (7 243)		*ac122* (62)	98·4 (61/62)		0·073
116	*pk-2*	100825←101472	215 (25 054)	E, L	*ac123* (215)	92·1 (198/215)	E, L	0·305
117		101653→102396	247 (28 704)	E, L	*ac124* (247)	96·4 (238/247)	E, L	0·168
118	*lef-7*	102414←103094	226 (26 549)		*ac125* (226)	91·2 (206/226)		0·623
119	*chitinase*	103143←104798	551 (61 701)	L	*ac126* (551)	96·6 (532/551)	L	0·224
120	*v-cath*	104844→105815	323 (36 964)	L	*ac127* (323)	99·1 (320/323)	E, L	0·032
121	*gp67*	106034←107623	529 (60 629)	E, L	*ac128* (530)	97·9 (519/530)	E, L	0·053
122	*p24capsid*	107752→108348	198 (22 109)	L	*ac129* (198)	93·4 (185/198)	L	0·234
123	*gp16*	108376→108696	106 (12 143)	L	*ac130* (106)	99·1 (105/106)	L	0·026
124	*pep*	108754→109719	321 (36 410)		*ac131* (322)	98·4 (317/322)	E	0·025
125		109722→110393	223 (25 692)	L	*ac132* (219)	90·6 (202/223)	L	0·370
126	*alk-exo*	110421→111683	420 (48 462)	L	*ac133* (419)	97·4 (409/420)	L	0·063
127		111743←114154	803 (94 645)		*ac134* (803)	95·5 (764/800)		0·242
	hr5	115321–115714						
128	*p35*	114362→115264	300 (35 137)	C, L	*ac135* (299)	94·0 (282/300)	C, L	0·187
129	*p26*	115771→116493	240 (27 372)		*ac136* (240)	92·3 (231/240)		0·165
130	*p10*	116567→116851	94 (10 296)	L	*ac137* (94)	94·7 (89/94)	L	0·136
131	*p74*	116861←118798	645 (73 887)	L	*ac138* (645)	97·8 (631/645)	L	0·067
132	*me53*	118908←120260	450 (52 745)	E, C, L	*ac139* (449)	95·3 (430/451)	E, C, L	0·133
132a		120269←120610	113 (12 733)		*ac139a* (67)	95·3 (61/64)		1·89
133	*ie-0*	120538→121323	261 (30 136)	C, L	*ac141* (261)	98·5 (257/261)	C, L	0·115
134		121338→122774	478 (55 541)	E, L	*ac142* (477)	99·4 (472/475)	E, L	0·015
135	*odv-e18*	122776→123048	90 (9 708)	L	*ac143* (90)	90·0 (81/90)	L	0·182
136	*odv-ec27*	123065→123937	290 (33 527)	L	*ac144* (290)	100·0 (290/290)	L	NC
137		123952→124239	95 (11 055)	L	*ac145* (95)	96·8 (92/95)	L	0·068
138		124234←124839	201 (22 807)		*ac146* (201)	98·0 (197/201)		0·065

Table 1. (*cont.*)

ORF	Name	Position*	aa (Da)†	Promoter motifs‡	Comparison with AcMNPV ORFs: AcMNPV homologue (aa)	% identity (range of alignment)	Promoter motifs‡	ω (d_N/d_S)§
139	*ie-1*	124905→126647	580 (66 691)	C	*ac147* (582)	97·1 (566/583)	C	0·058
140	*odv-e56*	126682←127818	378 (40 966)	L	*ac148* (376)	96·3 (364/378)	L	0·140
141		127847←128170	107 (12 341)		*ac149* (107)	94·4 (101/107)		0·319
142		128136→128429	97 (11 079)		*ac150* (99)	84·8 (84/99)		0·581
143	*ie-2*	128477←129700	407 (46 664)	E, C, L	*ac151* (408)	91·5 (376/411)	E, C, L	0·183
144		129726←130004	92 (10 872)		*ac152* (92)	88·0 (81/92)		0·984
145	*pe38*	130142→131113	323 (37 408)	E, C, L	*ac153* (321)	89·2 (290/325)	E, C, L	0·322
146		131233→131475	80 (9 323)	C, L	*ac154* (81)	92·6 (75/81)	C, L	0·276
	hr1	131515–326						

*Direction of ORF in the RoMNPV genome is indicated by an arrow.

†Number of amino acids encoded by ORF and molecular mass in Daltons.

‡Promoter motifs present upstream of ORF. C, Cap site CA(G/T)T 120 bp upstream of start codon, preceded by a TATA box TATA(A/T)A(A/T); E, early promoter motif CGTGC, 210 bp upstream of start codon; L, late promoter motif (A/T/G)TAAG 120 bp upstream of start codon.

§ω values (ratios of non-synonymous to synonymous mutation rates, averaged over the entire sequence) calculated by the pairwise method of the CODEML program of PAML (Yang, 1997). NC, not calculated due to the absence of non-synonymous or synonymous mutations.

from the RoMNPV genome due to mutations that reduce the size of the ORFs below 50 codons (Fig. 4a). For *ac97*, a G→A substitution in the RoMNPV homologue results in the appearance of a stop codon, terminating the ORF after three codons. The region containing *ac97* is missing from the BmNPV genome sequence (Gomi *et al.*, 1999). For *ac121* and *ac140*, single nucleotide insertions in the RoMNPV homologues result in frameshifts that lead to premature stop codons. ORFs *ac97*, *ac121* and *ac140* are not present in the genomes of any other lepidopteran NPVs and GVs. The presence of *ac97*, *ac121* and *ac140* in AcMNPV-C6 was confirmed by amplifying and sequencing the regions of these ORFs from our laboratory stock of AcMNPV-C6. Because of the small size of these ORFs and their truncation in RoMNPV, these ORFs may not be expressed in AcMNPV.

Forty-eight ORFs in RoMNPV differ in size from their AcMNPV homologues as a consequence of insertions, deletions and changes in stop codons. Of these, 26 ORFs differ in length by a single codon. Some RoMNPV ORFs are altered significantly in size (Fig. 4b). In *ro10*/*ac12*, a 5 nt insertion of 16 codons into the RoMNPV ORF causes a frameshift leading to a stop codon. An additional initiation codon occurs downstream and, after four codons, a single nucleotide deletion restores the original frame, resulting in an *ac12* homologue that is shortened by 25 codons. With *ro81*/*ac84*, a G→A substitution in the start codon results in an N-terminal truncation of the ORF by 24 codons. With *ro101*/*ac105*, deletions and insertions result in N-terminal extensions of 19 codons. Four additional ORFs (*ro15*/*ac17*, *ro49*/*ac52*, *ro135*/*ac143* and *ro137*/*ac145*) were determined

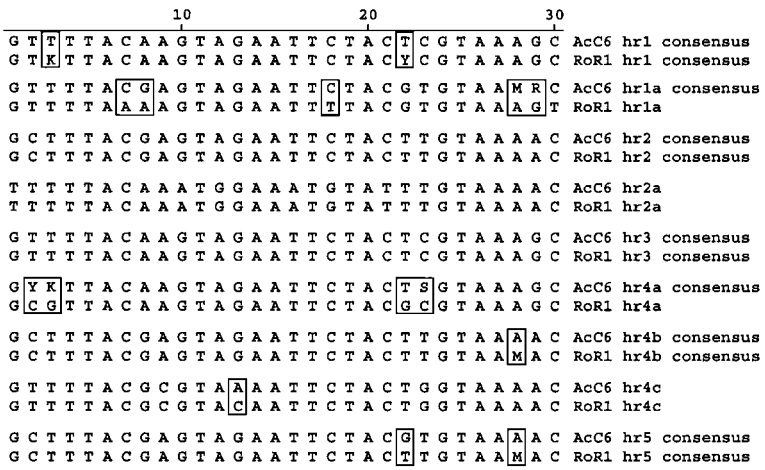

Fig. 2. Alignment of the consensus sequences of the 30 bp palindromic repeats for each *hr* of RoMNPV-R1 (RoR1) and AcMNPV-C6 (AcC6). Boxes indicate sequence mismatches. DNA codes: K = G or T; R = G or A; M = A or C; and Y = C or T.

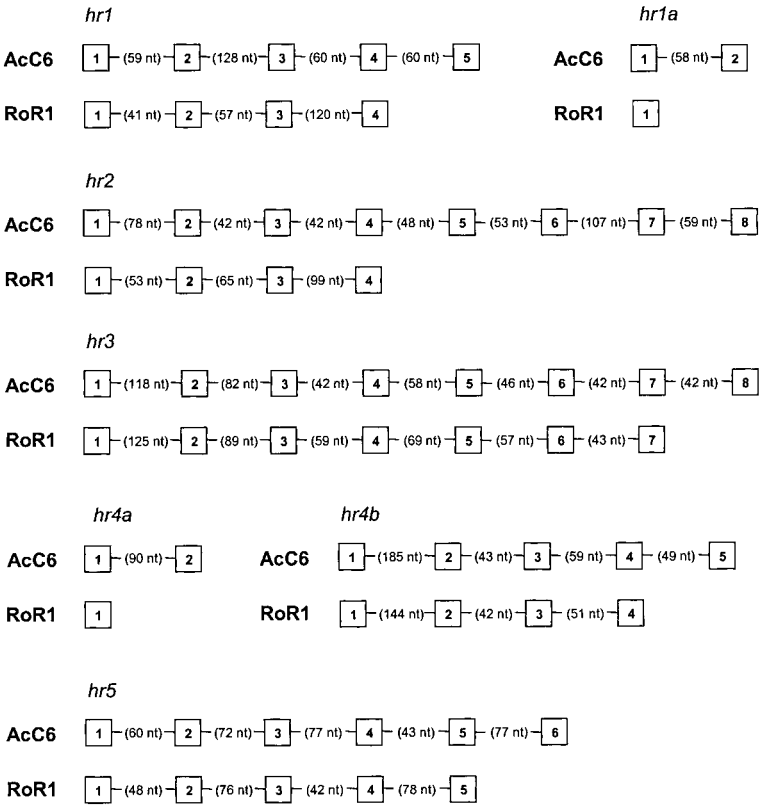

Fig. 3. Structure of AcMNPV-C6 (AcC6) and RoMNPV-R1 (RoR1) *hr*s showing the number of 30 bp palindromic repeats (open boxes) in each *hr*. The spacing between each repeat is indicated in nucleotides.

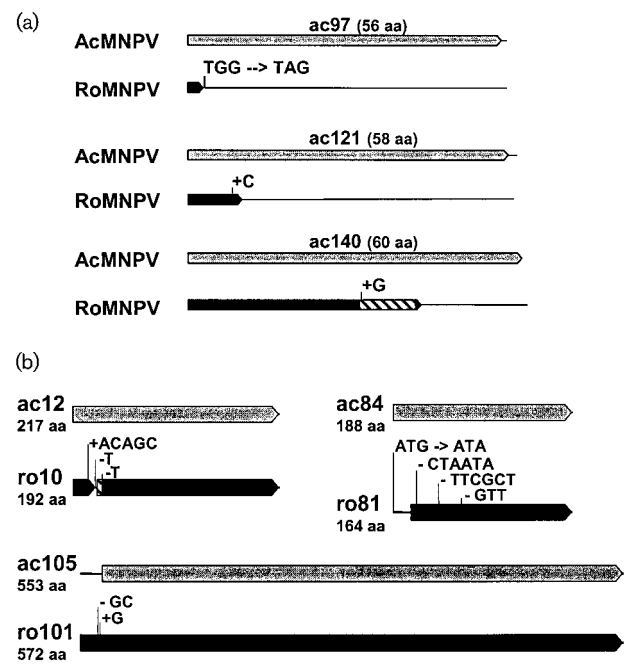

Fig. 4. (a) AcMNPV ORFs that are truncated in RoMNPV. (b) Comparison of AcMNPV and RoMNPV ORFs that differ significantly in size. Shaded (AcMNPV) and black (RoMNPV) arrows represent homologous ORFs. Substitutions, insertions and deletions in the RoMNPV ORFs resulting in their termination or expansion are indicated. The hatched region of *ro10* and the *ac140* homologue indicates non-homologous amino acid sequences that result from frameshifts.

initially to differ significantly in size, but PCR amplification and sequence analysis of these four ORFs from our laboratory stock of AcMNPV-C6 revealed that the AcMNPV ORFs were the same size as their RoMNPV homologues.

Four pairs of adjacent AcMNPV ORFs that are in the same orientation (*ac20*/*ac21*, *ac58*/*ac59*, *ac106*/*ac107* and *ac112*/*ac113*) are fused into a single ORF in RoMNPV. Where these ORFs occur in other baculovirus genomes, they are also fused into a single ORF. To confirm that these ORF pairs exist as separate ORFs in AcMNPV-C6, the sequences containing these ORFs were amplified from our laboratory stock of AcMNPV-C6 and subjected to DNA sequence analysis. In all four cases, the ORF pairs occurred as a single ORF in our stock of AcMNPV-C6. With respect to *ac20*/*ac21*, this finding is consistent with the sequence results obtained by Roncarati & Knebel-Mörsdorf (1997). It is not clear if these or other differences between the original AcMNPV-C6 genome sequence and our re-determined sequences represent errors in the original sequence or sequence properties unique to the AcMNPV-C6 stock sequenced by Ayres *et al.* (1994).

Of the ORFs that RoMNPV and AcMNPV have in common, the average predicted amino acid sequence identity (with one SD) is $96{\cdot}1 \pm 3{\cdot}12$ %. Twelve RoMNPV ORFs (*v-ubi*, *ro54*, *ro57*, *ro72*, *ro73*, *ro82*, *vp15*, *cg30*, *ro89*, *p6·9*, *ro96a* and *odv-ec27*) are completely identical in amino acid sequence (100 %) with their AcMNPV homologues along their entire length (Table 1). The most divergent ORF is *hcf-1* (host cell factor-1), with an amino acid sequence identity of

84·1 % between the AcMNPV and RoMNPV homologues. Mutations to eliminate expression of *hcf-1* during AcMNPV infection resulted in impairment of virus replication in two cells lines derived from *Trichoplusia ni* but not in a Sf cell line (Lu & Miller, 1996). *hcf-1* mutants killed *T. ni* more slowly. A reduction in the infectivity of *hcf-1* mutant BV, but not virus occlusions, towards *T. ni* larvae was also seen. In contrast, mutations in *hcf-1* had no effect on the dose of virus or the time required to kill *S. frugiperda* larvae (Lu & Miller, 1996). The relatively large degree of sequence divergence between AcMNPV and RoMNPV *hcf-1* and the species-specific effects of *hcf-1* mutation suggest that *hcf-1* may account for the different host range and virulence characteristics of AcMNPV and RoMNPV. However, *hcf-1* is required for optimal replication in *T. ni*, a species that is equally susceptible to both AcMNPV and AfMNPV, but not in *S. frugiperda*, which is more susceptible to AfMNPV than AcMNPV. It is not known if *hcf-1* influences virus replication in species other than *T. ni*.

All RoMNPV ORFs, except for *ro6* (polyhedrin, *polh*) and *ro76*, possess the greatest degree of amino acid sequence identity with AcMNPV homologues. The *ro76* ORF shows 96·2 % amino acid sequence identity with the BmNPV homologue *bm65* and 95·2 % identity with *ac79*. RoMNPV polyhedrin shows the greatest degree of amino acid sequence identity (98 %) with the predicted polyhedrin sequence of *Thysanoplusia orichalcea* MNPV. To examine the relationships of AcMNPV and RoMNPV polyhedrins to other baculovirus polyhedrins, phylogenetic trees of polyhedrin amino acid sequences were produced by two different methods. Both trees place AcMNPV polyhedrin on a branch outside of the clade containing the other group 1 NPVs (Fig. 5). The RoMNPV polyhedrin is found among group 1 NPV polyhedrins with a clade containing the polyhedrins from NPVs of *Thysanoplusia orichalcea*, *Antheraea pernyi* and *Attacus ricini*. In contrast, phylograms of NPV *dnapol* and *p10* predicted amino acid sequences place the AcMNPV and RoMNPV sequences together within the group 1 NPV clade (Fig. 6). These analyses suggest that AcMNPV acquired its *polh* gene by recombination with another virus that is not closely related to other group 1 NPVs.

A comparison of promoter motifs upstream of RoMNPV and AcMNPV ORFs (Table 1) revealed differences in the presence of early and late gene promoter motifs in 23 ORFs. Differences in promoter element composition may lead to alterations in the timing or level of transcription of these ORFs.

Analysis of selection pressure on RoMNPV and AcMNPV genes

Selection pressure analysis of vertebrate virus genes has identified positively selected sites that map to regions involved in host immune recognition and receptor binding (Woelk & Holmes, 2001; Woelk *et al.*, 2001; Holmes *et al.*, 2002; Twiddy *et al.*, 2002). Analysis of selection pressure on viral genes can potentially identify genes involved in virulence or in crossing of species barriers, even without prior knowledge of the mechanisms governing host range and virulence.

To detect instances of positive selection among AcMNPV and RoMNPV genes, ω was calculated for all of the RoMNPV and AcMNPV ORFs listed in Table 1 using the pairwise method available in PAML (Yang, 1997). The average value of ω from this analysis was 0·23, suggesting that, in general, negative selection pressure has been the dominant force in the evolution of most of the genes in the lineage containing RoMNPV and AcMNPV. Three ORFs were found to possess an ω value greater than 1: *ro65*/*ac68* ($\omega = 1{\cdot}20$), *ro110*/*ac116* ($\omega = 3{\cdot}50$) and *ro132a*/*ac139a* ($\omega = 1{\cdot}89$). Because function and protein expression has not been demonstrated for these ORFs, their values of d_N and d_S may be the product of random drift rather than selection at the protein level. The *ro110*/*ac116* and *ro132a*/*ac139a* ORFs have not been found in other baculovirus genomes. In contrast, *ro65*/*ac68* is present in all NPV and GV genomes sequenced previously, suggesting that its gene product plays an important role in the baculovirus life cycle. The value of ω for this ORF suggests that it is subject to a slight degree of positive selection pressure.

The pairwise method used to obtain the ω values shown in Table 1 calculates an average ω value for the entire ORF. Because a large number of amino acids in a protein are invariant ($\omega = 0$) due to functional constraints, it is difficult to detect positive selection using this approach. To overcome this problem, a selection of RoMNPV and AcMNPV ORFs was analysed with models that allow for heterogeneous ω values among different codon sites (Yang *et al.*, 2000; see Methods). The ORFs analysed consisted of 63 genes present in nine other baculovirus genomes (Herniou *et al.*, 2001) and other ORFs for which protein expression or functional activity had been demonstrated for AcMNPV. Models M2, M3 and M7 identified positively selected sites in several ORFs. However, the null hypothesis models (M0, M1 and/or M7) could be rejected at the $P < 0{\cdot}05$ level only for *ro18*/*ac20-ac21* (*arif-1*) and *ro135*/*ac143* (*odv-e18*) (Table 2). Re-sequencing of the *ac20-ac21* and *ac143* ORFs confirmed that the results from analysis of selection pressures on these genes were not based on AcMNPV sequences containing potential errors. The *arif-1* ORF encodes the 48 kDa actin rearrangement-inducing factor, a protein that localizes to vesicular structures at the plasma membrane of infected cells (Roncarati & Knebel-Mörsdorf, 1997). ARIF-1 mediates the dissociation of the host cell actin network and of virus-induced actin cables that form early during infection, as well as subsequent formation of actin aggregates at the plasma membrane (Dreschers *et al.*, 2001). Mutations in AcMNPV *arif-1* had no effect upon replication in *S. frugiperda* or *T. ni* cells *in vitro* (Roncarati & Knebel-Mörsdorf, 1997; Dreschers *et al.*, 2001). ODV-E18 migrates as a potential dimer on protein gels and is located in

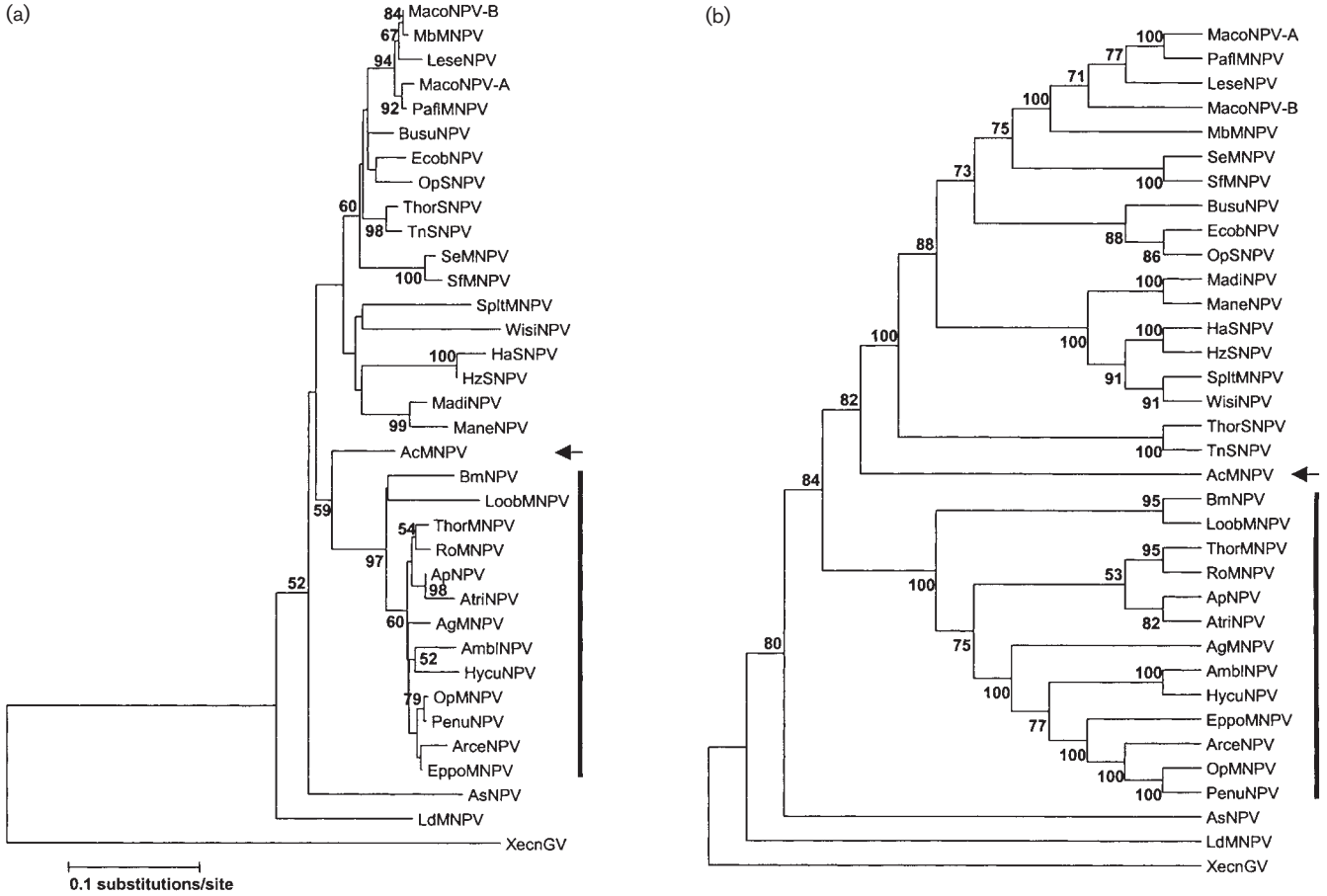

Fig. 5. Phylogenetic analysis of amino acid sequences from 34 polyhedrin genes and 1 granulin gene. (a) ME phylogram. Bootstrap values $\geqslant 50\,\%$ ($n=1000$ replicates) are shown at interior branches where they occur. (b) MP cladogram. The topology of the tree is a majority consensus of 45 equally parsimonious trees. The percentage of trees displaying the given branching pattern at each node is shown. AcMNPV, *Autographa californica* MNPV (Ayres *et al.*, 1994); AgMNPV, *Anticarsia gemmatalis* MNPV (Zanotto *et al.*, 1992); AmblNPV, *Amsacta albistriga* NPV (accession no. AF118850); ApNPV, *Antheraea pernyi* NPV (accession no. AB062454); ArceNPV, *Archips cerasivoranus* NPV (accession no. U40834); AsNPV, *Agrotis segetum* NPV (Kozlov *et al.*, 1992); AtriNPV, *Attacus ricini* NPV (accession no. S68462); BmNPV, *Bombyx mori* NPV (Gomi *et al.*, 1999); BusuNPV, *Buzuria suppressaria* NPV (Hu *et al.*, 1993); EcobNPV, *Ecotropis oliqua* NPV (accession no. U95014); EppoMNPV, *Epiphyas postvittana* MNPV (Hyink *et al.*, 2002); HaSNPV, *Helicoverpa armigera* SNPV (Chen *et al.*, 2001); HzSNPV, *Helicoverpa zea* SNPV (Chen *et al.*, 2002); HycuNPV, *Hyphantria cunea* NPV (accession no. D14573); LdMNPV, *Lymantria dispar* MNPV (Kuzio *et al.*, 1999); LeseNPV, *Leucania separata* NPV (accession no. AAB47865); LoobMNPV, *Lonomia obliqua* MNPV (accession no. AAF98122); MacoNPV-A, *Mamestra configurata* NPV strain 90/2 (Li *et al.*, 2002b); MacoNPV-B, *Mamestra configurata* NPV strain 96B (Li *et al.*, 2002a); MadiNPV, *Malacosoma disstria* NPV (accession no. AAD00095); ManeNPV, *Malacosoma neustria* NPV (accession no. AAB31529); MbMNPV, *Mamestra brassicae* MNPV (Cameron & Possee, 1989); OpMNPV, *Orgyia pseudotsugata* MNPV (Ahrens *et al.*, 1997); OpSNPV, *Orgyia pseudotsugata* SNPV (Leisy *et al.*, 1986); PaflMNPV, *Panolis flammea* MNPV (Oakey *et al.*, 1989); PenuNPV, *Perina nuda* NPV (Chou *et al.*, 1996); RoMNPV, *Rachiplusia ou* MNPV (Harrison & Bonning, 1999); SeMNPV, *Spodoptera exigua* MNPV (IJkel *et al.*, 1999); SfMNPV, *Spodoptera frugiperda* MNPV (Gonzalez *et al.*, 1989); SpltMNPV, *Spodoptera litura* MNPV (Pang *et al.*, 2001); ThorMNPV, *Thysanoplusia orichalcea* MNPV (accession no. AF169480); ThorSNPV, *Thysanoplusia orichalcea* SNPV (Cheng *et al.*, 1998); TnSNPV, *Trichoplusia ni* SNPV (Fielding & Davison, 1999); WisiNPV, *Wiseana signata* NPV (accession no. AF016916); XecnGV, *Xestia c-nigrum* granulovirus (Hayakawa *et al.*, 1999).

virus-associated intranuclear membranes and envelopes of occlusion-derived virus (Braunagel *et al.*, 1996). The *odv-e18* ORF also appears to encode the N-terminal portion of a larger occluded virus envelope protein, ODV-E35 (Braunagel *et al.*, 1996).

For *arif-1*, models M2 and M3 identified categories of codons consisting of approximately 1 % of all codon sites and possessing very high ω values (ω = 80·55 for M2 and ω = 59·44 for M3; Table 2), indicating a strong degree of positive selection. In likelihood ratio tests, the M0 and M1

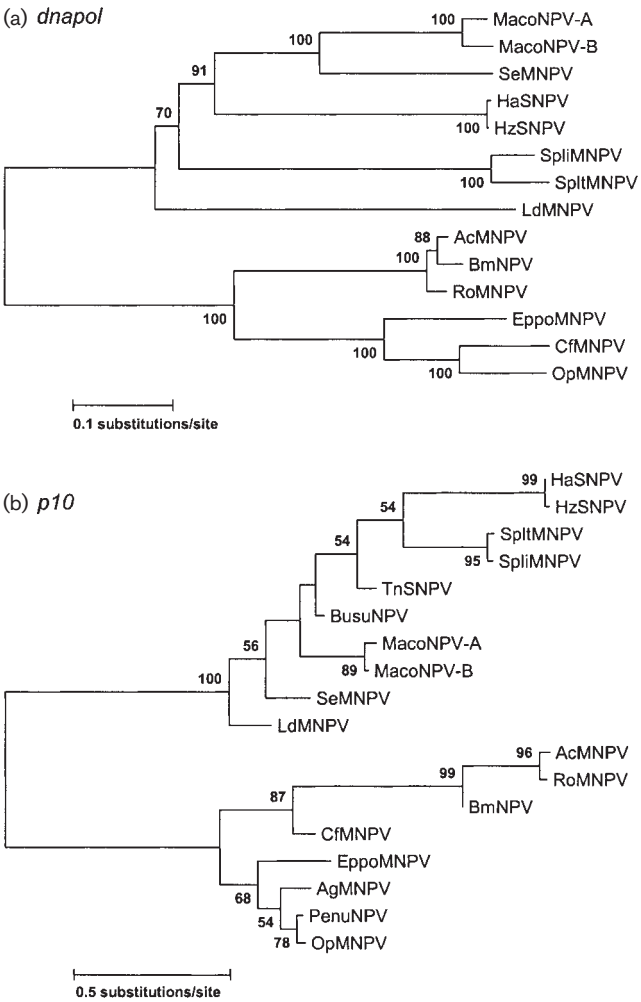

Fig. 6. Phylogenetic analysis of predicted amino acid sequences of NPV *dnapol* (a) and *p10* (b) genes. Both trees are phylograms produced by ME with bootstrap values $\geqslant 50\,\%$ ($n=1000$ replicates) displayed at interior branches where they occur. NPV abbreviations are as indicated for Fig. 5, with the addition of CfMNPV (*Choristoneura fumiferana* MNPV). The sources of *dnapol* and *p10* sequences for AcMNPV, BmNPV, EppoMNPV, HaSNPV, HzSNPV, LdMNPV, MacoNPV-A, MacoNPV-B, OpMNPV, SeMNPV and SpltMNPV are as described for Fig. 5. CfMNPV *dnapol*, Liu & Carstens (1995); SpliMNPV *dnapol*, accession no. AF215639; AgMNPV *p10*, accession no. AY055828; BusuNPV *p10*, van Oers *et al.* (1998); CfMNPV *p10*, Wilson *et al.* (1995); PenuNPV *p10*, accession no. U50411; SpliMNPV *p10*, Faktor *et al.* (1997); TnSNPV *p10*, accession no. AF358416.

models could be rejected in favour of the M2 model at $P<0{\cdot}05$, but only the M0 model could be rejected in favour of the M3 model. Model M8 identified a positively selected codon category consisting of a larger number of sites (9·6 % of all codons) but possessing a much lower ω value (3·130). Null hypothesis model M7 could not be rejected in favour of M8 ($P=0{\cdot}336$). Models M2 and M3 identified codon positions 94, 365 and 390 in *arif-1* as being in the positively selected category, although only positions 365 and 390 were identified with a probability greater than 0·90 in both models. For all three sites, substitutions had taken place at two positions within the codons with at least one transversion. Positions 365 and 390 are located in a region of the protein localized to the cytoplasmic side of the plasma membrane (Dreschers *et al.*, 2001).

In addition to the cap site and TATA box in the upstream regions of both AcMNPV and RoMNPV *arif-1*, the RoMNPV gene also contains a copy of the late gene promoter motif. It is unclear if the potential late phase transcription of RoMNPV *arif-1* would have any impact on the function of ARIF-1 during infection that would influence its host range. AcMNPV ARIF-1 protein is detectable as late as 48 h p.i, but after 12 h p.i., it is phosphorylated and found solely in cytoplasmic vacuoles. Dreschers *et al.* (2001) speculate that ARIF-1 is rendered non-functional by phosphorylation. Also, the formation of filamentous actin in the nucleus of infected cells that takes place during the late phase of infection does not appear to involve ARIF-1 (Roncarati & Knebel-Mörsdorf, 1997; Ohkawa *et al.*, 2002). Although mutations in *arif-1* had no effect upon infection and replication *in vitro*, ARIF-1 may be required for the *in vivo* replication cycle.

The evidence for positive selection of *odv-e18* was much stronger than for *arif-1*. Models M2, M3 and M8 all identified positively selected codon categories consisting of the same 10 sites (positions 52, 55–61, 65 and 66) at *P* values well below 0·05. ω values were very high for all three models (99·00), indicating a very strong degree of positive selection. Bayesian analysis yielded probabilities greater than 0·95 for all 10 sites, except position 55 in model M3. Seven of the positively selected sites occurred contiguously in a region of high nucleotide and amino acid sequence divergence. Oddly, one of the positively selected sites (position 56) encodes a serine in both AcMNPV and RoMNPV genes. The codon positions at this site (TCG in *ac143* and AGC in *ro135*) differ by transversions at every position. The codon-based substitution models implemented in PAML assume that only one codon position changes at a time (Yang, 2001). To change from TCG to AGC, one position at a time, at least two non-synonymous substitutions must first occur, which may account for why position 56 was identified as a positively selected site. Because of its location, ODV-E18 may influence host range at the level of midgut cell binding and internalization of occluded virus.

Models M0, M2, M3 and M8 all calculated an ω of 1·14 for *ro65*/*ac68*, indicating a weak degree of positive selection for this ORF, but the null hypothesis models could not be rejected at $P<0{\cdot}05$. Hence, the inference of selection pressure on this ORF should be treated with caution.

The power of the likelihood ratio test is defined as the probability of rejecting the null hypothesis when it is wrong and when the alternative hypothesis (in this case, the inference of positive selection) is correct. This probability

Table 2. Positive selection analysis of *ro18/ac20-ac21* and *ro135/ac143*

Model	ln*L**	Average d_N/d_S	Parameter estimates†	Likelihood ratio tests‡	Positively selected sites (probability)§
***ro18/ac20-ac21* (*arif-1*)**					
M0	−1 927·512	0·207	$\omega = 0{\cdot}207$		
M1	−1 924·356	0·209	$p_0 = 0{\cdot}792$		
			$p_1 = 0{\cdot}208$		
M2	−1 921·178	0·922	$p_0 = 0{\cdot}799$	M0 versus M2:	A94L (0·7804)
			$p_1 = 0{\cdot}192$	$P = 0{\cdot}002$	K365V (0·9281)
			$\mathbf{p_2 = 0{\cdot}009}$	M1 versus M2:	P390R (0·9548)
			$\mathbf{\omega_2 = 80{\cdot}55}$	$P = 0{\cdot}042$	
M3	−1 920·989	0·834	$p_0 = 0{\cdot}000$	M0 versus M3:	P55L (0·5354)
			$p_1 = 0{\cdot}989$	$P = 0{\cdot}011$	A94L (0·9710)
			$\mathbf{p_2 = 0{\cdot}011}$	M1 versus M3:	K365V (0·9919)
			$\omega_0 = 0{\cdot}000$	$P = 0{\cdot}151$	P390R (0·9560)
			$\omega_1 = 0{\cdot}159$		
			$\mathbf{\omega_2 = 59{\cdot}44}$		
M7	−1 924·370	0·200	$p = 0{\cdot}001$		
			$q = 0{\cdot}005$		
M8	−1 923·279	0·300	$p = 0{\cdot}001$	M7 versus M8: 0·336	
			$q = 1{\cdot}68$		
			$p_0 = 0{\cdot}904$		
			$\mathbf{p_1 = 0{\cdot}096}$		
			$\mathbf{\omega_1 = 3{\cdot}130}$		
***ro135/ac143* (*odv-e18*)**					
M0	−495·881	0·180	$\omega = 0{\cdot}180$		
M1	−483·152	0·164	$p_0 = 0{\cdot}836$		
			$p_1 = 0{\cdot}164$		
M2	−472·067	11·94	$p_0 = 0{\cdot}879$	M0 versus M2:	N52T (1·00)
			$p_1 = 0{\cdot}000$	$P < 10^{-7}$	S55N (1·00)
			$\mathbf{p_2 = 0{\cdot}121}$	M1 versus M2:	S56S (1·00)
			$\mathbf{\omega_2 = 99{\cdot}00}$	$P = 1{\cdot}5 \times 10^{-5}$	G57S (1·00)
					G58S (1·00)
					N59P (1·00)
					V60N (1·00)
					P61T (1·00)
					L65N (1·00)
					G66M (1·00)
M3	−472·052	11·70	$p_0 = 0{\cdot}000$	M0 versus M3:	N52T (0·9785)
			$p_1 = 0{\cdot}882$	$P < 10^{-7}$	S55N (0·8244)
			$\mathbf{p_2 = 0{\cdot}118}$	M1 versus M3:	S56S (1·00)
			$\omega_1 = 0{\cdot}000$	$P = 1{\cdot}83 \times 10^{-4}$	G57S (0·9963)
			$\omega_2 = 0{\cdot}004$		G58S (1·00)
			$\mathbf{\omega_3 = 99{\cdot}00}$		N59P (1·00)
					V60N (1·00)
					P61T (0·9925)
					L65N (1·00)
					G66M (1·00)
M7	−483·491	0·201	$p = 0{\cdot}028$		
			$q = 0{\cdot}152$		
M8	−472·067	11·94	$p = 0{\cdot}001$	M7 versus M8:	N52T (1·00)
			$q = 1{\cdot}817$	$1{\cdot}1 \times 10^{-5}$	S55N (1·00)
			$p_0 = 0{\cdot}879$		S56S (1·00)
			$\mathbf{p_1 = 0{\cdot}121}$		G57S (1·00)
			$\mathbf{\omega_1 = 99{\cdot}00}$		G58S (1·00)
					N59P (1·00)
					V60N (1·00)

Table 2. (*cont.*)

Model	ln*L**	Average d_N/d_S	Parameter estimates†	Likelihood ratio tests‡	Positively selected sites (probability)§
					P61T (1·00) L65N (1·00) G66M (1·00)

*Log-likelihood value.

†Numerical values for site categories (p_0, p_1 and p_2) indicate the proportion of codons in each category. p and q are parameters that determine the shape of the β distribution of ω values in models M7 and M8. Parameters indicating positive selection are in bold type.

‡P value determined by comparing twice the difference of ln$L$ of the models with a χ^2 distribution with degrees of freedom equal to 2 (for M0 versus M2, M1 versus M2 and M7 versus M8) or 4 (for M0 versus M3 and M1 versus M3).

§Positively selected sites are indicated by their position numbers, preceded by the identity of the amino acid in the AcMNPV ORF and followed by the identity of the amino acid in the RoMNPV ORF.

decreases with decreasing number of sequences per data set but increases with increasing strength of positive selection (Anisimova *et al.*, 2001). Hence, it is not surprising that, with data sets consisting of only two sequences (one RoMNPV and one AcMNPV gene), we were only able to reject the null hypothesis models in two instances where the strength of selection (the value of ω) was very high. Selection pressure analysis that includes sequences from other NPV genomes may identify other genes undergoing positive selection.

Because of the genetic similarity between AcMNPV and RoMNPV, study of these viruses may enhance our knowledge of the genetic bases of baculovirus host range. In addition to genes identified as undergoing positive selection, ORFs that differ in size, exhibit a relatively large degree of amino acid sequence divergence or show differences in timing or level of gene expression due to differences in promoter organization may also play a role in host range and virulence. This comparison of the genomes of RoMNPV and AcMNPV will serve as the foundation for empirical study of the molecular basis of host range and virulence differences between these two viruses.

ACKNOWLEDGEMENTS

We thank Michael Hensel for assistance with cloning restriction fragments, Ziheng Yang (Department of Biology, University College London, UK) and Gavin Naylor (Department of Zoology and Genetics, Iowa State University, USA) for helpful discussions, and Gary Polking and John Mlocek of the Iowa State University DNA Synthesis and Sequencing Facility. This study of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa, Project No. 3301 was supported by Hatch Act and State of Iowa funds.

REFERENCES

Adams, J. R. & McClintock, J. T. (1991). *Baculoviridae.* Nuclear polyhedrosis viruses. I. Nuclear polyhedrosis viruses of insects. In *Atlas of Invertebrate Viruses*, pp. 87–204. Edited by J. R. Adams & J. R. Bonami. Boca Raton, FL: CRC Press.

Ahrens, C. H., Russell, R. L. Q., Funk, C. J., Evans, J. T., Harwood, S. H. & Rohrmann, G. F. (1997). The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. *Virology* **229**, 381–399.

Anisimova, M., Bielawski, J. P. & Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol* **18**, 1585–1592.

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., Keddie, B. A., Possee, R., Rohrmann, G., Theilmann, D. & Volkman, L. (2000). Family *Baculoviridae*. In *Virus Taxonomy. Seventh Report of the International Committee on the Taxonomy of Viruses*, pp. 195–202. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle & R. B. Wickner. Vienna/New York: Springer-Verlag.

Braunagel, S. C., He, H., Ramamurthy, P. & Summers, M. D. (1996). Transcription, translation, and cellular localization of three *Autographa californica* nuclear polyhedrosis virus structural proteins: ODV-E18, ODV-E35, and ODV-EC27. *Virology* **222**, 100–114.

Cameron, I. R. & Possee, R. D. (1989). Conservation of polyhedrin gene promoter function between *Autographa californica* and *Mamestra brassicae* nuclear polyhedrosis viruses. *Virus Res* **12**, 183–199.

Cardenas, F. A., Vail, P. V., Hoffmann, D. F., Tebbets, J. S. & Schreiber, F. E. (1997). Infectivity of celery looper (Lepidoptera: Noctuidae) multiple nucleocapsid polyhedrosis virus to navel orangeworm (Lepidoptera: Pyralidae). *Environ Entomol* **26**, 131–134.

Chen, X., IJkel, W. F., Tarchini, R. & 8 other authors (2001). The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. *J Gen Virol* **82**, 241–257.

Chen, X., Zhang, W.-J., Wong, J. & 9 other authors (2002). Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. *J Gen Virol* **83**, 673–684.

Cheng, X. W., Carner, G. R. & Fescemyer, H. W. (1998). Polyhedrin sequence determines the tetrahedral shape of occlusion bodies in *Thysanoplusia orichalcea* single-nucleocapsid nucleopolyhedrovirus. *J Gen Virol* **79**, 2549–2556.

Chou, C. M., Huang, C. J., Lo, C. F., Kou, G. H. & Wang, C. H. (1996). Characterization of *Perina nuda* nucleopolyhedrovirus (PenuNPV) polyhedrin gene. *J Invertebr Pathol* **67**, 259–266.

Cunningham, J. C. (1995). Baculoviruses as microbial insecticides. In *Novel Approaches to Integrated Pest Management*, pp. 261–292. Edited by R. Reuveni. Boca Raton, FL: CRC Press.

Dreschers, S., Roncarati, R. & Knebel-Mörsdorf, D. (2001). Actin rearrangement-inducing factor of baculoviruses is tyrosine phosphorylated and colocalizes to F-actin at the plasma membrane. *J Virol* **75**, 3771–3778.

Eldridge, R., Li, Y. & Miller, L. K. (1992). Characterization of a baculovirus gene encoding a small conotoxinlike polypeptide. *J Virol* **66**, 6563–6571.

Faktor, O., Toister-Achituv, M., Nahum, O. & Kamensky, B. (1997). The *p10* gene of *Spodoptera littoralis* nucleopolyhedrovirus: nucleotide sequence, transcriptional analysis and unique gene organization in the *p10* locus. *J Gen Virol* **78**, 2119–2128.

Federici, B. A. & Hice, R. H. (1997). Organization and molecular characterization of genes in the polyhedrin region of the *Anagrapha falcifera* multinucleocapsid NPV. *Arch Virol* **142**, 333–348.

Fielding, B. C. & Davison, S. (1999). The characterization and phylogenetic relationship of the *Trichoplusia ni* single capsid nuclear polyhedrosis virus polyhedrin gene. *Virus Genes* **19**, 67–72.

Gomi, S., Majima, K. & Maeda, S. (1999). Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *J Gen Virol* **80**, 1323–1337.

Gonzalez, M. A., Smith, G. E. & Summers, M. D. (1989). Insertion of the SfMNPV polyhedrin gene into an AcMNPV polyhedrin deletion mutant during viral infection. *Virology* **170**, 160–175.

Granados, R. R. & Williams, K. A. (1986). *In vivo* infection and replication of baculoviruses. In *The Biology of Baculoviruses*, pp. 89–108. Edited by R. R. Granados & B. A. Federici. Boca Raton, FL: CRC Press.

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.

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.-I., 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., Vlak, 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.

Holmes, E. C., Woelk, C. H., Kassis, R. & Bourhy, H. (2002). Genetic constraints and the adaptive evolution of rabies virus in nature. *Virology* **292**, 247–257.

Hostetter, D. L. & Puttler, B. (1991). A new broad host spectrum nuclear polyhedrosis virus isolated from a celery looper, *Anagrapha falcifera* (Kirby), (Lepidoptera: Noctuidae). *Environ Entomol* **20**, 1480–1488.

Hu, Z. H., Liu, M. F., Jin, F., Wang, Z. X., Liu, X. Y., Li, M. J., Liang, B. F. & Xie, T. E. (1993). Nucleotide sequence of the *Buzura suppressaria* single nucleocapsid nuclear polyhedrosis virus polyhedrin gene. *J Gen Virol* **74**, 1617–1620.

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. & Vlak, J. M. (1999). Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *J Gen Virol* **80**, 3289–3304.

Jewell, J. E. & Miller, L. K. (1980). DNA sequence homology relationships among six lepidopteran nuclear polyhedrosis viruses. *J Gen Virol* **48**, 161–175.

Kang, W., Suzuki, M., Zemskov, E., Okano, K. & Maeda, S. (1999). Characterization of baculovirus repeated open reading frames (*bro*) in *Bombyx mori* nucleopolyhedrovirus. *J Virol* **73**, 10339–10345.

Kozlov, E. A., Rodnin, N. V., Levitina, T. L., Gusak, N. M., Radomskij, N. F. & Palchikovskaya, L. J. (1992). The amino acid sequence determination of a granulin and polyhedrin from two baculoviruses infecting *Agrotis segetum*. *Virology* **189**, 320–323.

Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**, 1244–1245.

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.

Leisy, D., Nesson, M., Pearson, M., Rohrmann, G. & Beaudreau, G. (1986). Location and nucleotide sequence of the *Orgyia pseudotsugata* single nucleocapsid nuclear polyhedrosis virus polyhedrin gene. *J Gen Virol* **67**, 1073–1079.

Lewis, L. C. & Johnson, T. B. (1982). Efficacy of two nuclear polyhedrosis viruses against *Ostrinia nubilalis* (Lep: Pyralidae) in the laboratory and field. *Entomophaga* **27**, 33–38.

Li, L., Donly, C., Li, Q., Willis, L. G., Keddie, B. A., Erlandson, M. A. & Theilmann, D. A. (2002a). Identification and genomic analysis of a second species of nucleopolyhedrovirus isolated from *Mamestra configurata*. *Virology* **297**, 226–244.

Li, Q., Donly, C., Li, L., Willis, L. G., Theilmann, D. A. & Erlandson, M. (2002b). Sequence and organization of the *Mamestra configurata* nucleopolyhedrovirus genome. *Virology* **294**, 106–121.

Liu, J. J. & Carstens, E. B. (1995). Identification, localization, transcription, and sequence analysis of the *Choristoneura fumiferana* nuclear polyhedrosis virus DNA polymerase gene. *Virology* **209**, 538–549.

Lu, A. & Miller, L. K. (1996). Species-specific effects of the *hcf-1* gene on baculovirus virulence. *J Virol* **70**, 5123–5130.

Moscardi, F. (1999). Assessment of the application of baculoviruses for control of Lepidoptera. *Annu Rev Entomol* **44**, 257–289.

Nei, M. & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.

Oakey, R., Cameron, I. R., Davis, B., Davis, E. & Possee, R. D. (1989). Analysis of transcription initiation in the *Panolis flammea* nuclear polyhedrosis virus polyhedrin gene. *J Gen Virol* **70**, 769–775.

Ohkawa, T., Rowe, A. R. & Volkman, L. E. (2002). Identification of six *Autographa californica* multicapsid nucleopolyhedrovirus early genes that mediate nuclear localization of G-actin. *J Virol* **76**, 12281–12289.

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.

Paschke, J. D. & Hamm, J. J. (1961). A nuclear polyhedrosis of *Rachiplusia ou* (Guenée). *J Insect Pathol* **3**, 333–334.

Payne, C. C. (1986). Insect pathogenic viruses as pest control agents. In *Biological Plant and Health Protection*. Edited by J. M. Franz. Stuttgart: Fischer-Verlag.

Possee, R. D. & Rohrmann, G. F. (1997). Baculovirus genome organization and evolution. In *The Baculoviruses*, pp. 109–140. Edited by L. K. Miller. New York: Plenum.

Roncarati, R. & Knebel-Mörsdorf, D. (1997). Identification of the early actin-rearrangement-inducing factor gene, *arif-1*, from

Autographa californica multicapsid nuclear polyhedrosis virus. *J Virol* **71**, 7933–7941.

Smith, G. E. & Summers, M. D. (1980). Restriction map of *Rachiplusia ou* and *Rachiplusia ou-Autographa californica* baculovirus recombinants. *J Virol* **33**, 311–319.

Smith, G. E. & Summers, M. D. (1982). DNA homology among subgroup A, B, and C baculoviruses. *Virology* **123**, 393–406.

Smith, C. R., Heinz, K. M., Sansone, C. G. & Flexner, J. L. (2000). Impact of recombinant baculovirus applications on target heliothines and non-target predators in cotton. *Biol Control* **19**, 201–214.

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.

Treacy, M. F. & All, J. N. (1996). Impact of insect-specific AaHIT gene insertion on inherent bioactivity of baculovirus against tobacco budworm, *Heliothis virescens*, and cabbage looper, *Trichoplusia ni*. In *Beltwide Cotton Conference*, pp. 911–917. Nashville, TN, USA.

Treacy, M. F., Rensner, P. E. & All, J. N. (2000). Comparative insecticidal properties of two nucleopolyhedrovirus vectors encoding a similar toxin gene chimer. *J Econ Entomol* **93**, 1096–1104.

Twiddy, S. S., Woelk, C. H. & Holmes, E. C. (2002). Phylogenetic evidence for adaptive evolution of dengue viruses in nature. *J Gen Virol* **83**, 1679–1689.

Vail, P. V., Hoffmann, D. F., Streett, D. A., Manning, J. S. & Tebbets, J. S. (1993). Infectivity of a nuclear polyhedrosis virus isolated from *Anagrapha falcifera* (Lepidoptera: Noctuidae) against production and postharvest pests and homologous cell lines. *Environ Entomol* **22**, 1140–1145.

van Oers, M. M., Hu, Z., Arif, B. M., van Strien, E. A., van Lent, J. W. & Vlak, J. M. (1998). The single-nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria* encodes a P10 protein. *J Gen Virol* **79**, 1553–1562.

Vaughn, J. L., Goodwin, R. H., Tompkins, G. J. & McCawley, P. (1977). The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera; Noctuidae). *In Vitro* **13**, 213–217.

Washburn, J. O., Haas-Stapleton, E. J., Tan, F. F., Beckage, N. E. & Volkman, L. E. (2000). Co-infection of *Manduca sexta* larvae with polydnavirus from *Cotesia congregata* increases susceptibility to fatal infection by *Autographa californica* M nucleopolyhedrovirus. *J Insect Physiol* **46**, 179–190.

Wilson, J. A., Hill, J. E., Kuzio, J. & Faulkner, P. (1995). Characterization of the baculovirus *Choristoneura fumiferana* multicapsid nuclear polyhedrosis virus *p10* gene indicates that the polypeptide contains a coiled-coil domain. *J Gen Virol* **76**, 2923–2932.

Woelk, C. H. & Holmes, E. C. (2001). Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). *J Mol Evol* **52**, 182–192.

Woelk, C. H., Jin, L., Holmes, E. C. & Brown, D. W. G. (2001). Immune and artificial selection in the haemagglutinin (H) glycoprotein of measles virus. *J Gen Virol* **82**, 2463–2474.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**, 555–556.

Yang, Z. (2001). Adaptive molecular evolution. In *Handbook of Statistical Genetics*, pp. 327–350. Edited by D. J. Balding, M. Bishop & C. Cannings. London: John Wiley & Sons.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A.-M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.

Zanotto, P. M., Sampaio, M. J., Johnson, D. W., Rocha, T. L. & Maruniak, J. E. (1992). The *Anticarsia gemmatalis* nuclear polyhedrosis virus polyhedrin gene region: sequence analysis, gene product and structural comparisons. *J Gen Virol* **73**, 1049–1056.

Zemskov, E. A., Kang, W. & Maeda, S. (2000). Evidence for nucleic acid binding ability and nucleosome association of *Bombyx mori* nucleopolyhedrovirus BRO proteins. *J Virol* **74**, 6784–6789.