Analysis of the *Choristoneura fumiferana* nucleopolyhedrovirus genome

Jondavid G. de Jong,[1] Hilary A. M. Lauzon,[2] Cliff Dominy,[1] Arkadi Poloumienko,[1] Eric B. Carstens,[3] Basil M. Arif[2] and Peter J. Krell[1]

[1]Department of Molecular and Cellular Biology, University of Guelph, Guelph, Ontario, Canada, N1G 2W1

[2]Great Lakes Forestry Centre, Sault Ste Marie, Ontario, Canada, P6A 2E5

[3]Department of Microbiology and Immunology, Queens University, Kingston, Ontario, Canada, K7L 3N6

Correspondence
Peter J. Krell
pkrell@uoguelph.ca

Received 30 July 2004
Accepted 12 November 2004

The double-stranded DNA genome of *Choristoneura fumiferana* nucleopolyhedrovirus (CfMNPV) was sequenced and analysed in the context of other group I nucleopolyhedroviruses (NPVs). The genome consists of 129 593 bp with a G+C content of 50·1 mol%. A total of 146 open reading frames (ORFs) of greater than 150 bp, and with no or minimal overlap were identified. In addition, five homologous regions were identified containing 7–10 repeats of a 36 bp imperfect palindromic core. Comparison with other completely sequenced baculovirus genomes revealed that 139 of the CfMNPV ORFs have homologues in at least one other baculovirus and seven ORFs are unique to CfMNPV. Of the 117 CfMNPV ORFs common to all group I NPVs, 12 are exclusive to group I NPVs. Overall, CfMNPV is most similar to *Orgyia pseudotsugata MNPV* based on gene content, arrangement and overall amino acid identity. Unlike other group I baculoviruses, however, CfMNPV encodes a viral enhancing factor (*vef*) and has two copies of *p26*.

INTRODUCTION

Baculoviridae is a family of large, enveloped viruses characterized by rod-shaped virions and covalently closed double-stranded circular DNA genomes ranging in size from 80 to 180 kb (Blissard *et al.*, 2000). *Baculoviridae* contains two genera; the *Nucleopolyhedrovirus* (NPVs) and *Granulovirus* (GVs) genera, based on their occlusion body morphology. Baculoviruses are infectious only to arthropods, mostly to insects within the orders *Lepidoptera*, *Diptera* or *Hymenoptera* (Miller, 1997). Baculoviruses have been evaluated and used as bio-insecticides, for the over-expression of foreign proteins, as vectors for expression in mammalian cells and in baculovirus surface display technology (Kost & Condreay, 1999).

The genome sequences of 22 baculoviruses are published and 26 are currently listed in GenBank. Herniou *et al.* (2003) identified 30 core baculovirus genes common to 13 analysed baculovirus genomes. With the addition of the complete genome sequences of *Neodiprion lecontei NPV* (NeleNPV; Lauzon *et al.*, 2004) and *Neodiprion sertifer NPV* (NeseNPV; Garcia-Maruniak *et al.*, 2004) the core set of baculoviral genes has dropped to 29 as both hymenopteran baculoviruses lacked an identified F protein. Based on phylogenetic analysis, using the 29 core genes, baculoviruses segregate into five major groups, the GVs, the group I and group II NPVs, a group of the dipteran virus, *Culex nigripalpus NPV* (CuniNPV; Afonso *et al.*, 2001) and a group containing the hymenopteran viruses. The five published GV genomes include those for *Xestia c-nigrum GV* (XecnGV; Hayakawa *et al.*, 1999), *Plutella xylostella GV* (PlxyGV; Hashimoto *et al.*, 2000), *Cydia pomonella GV* (CpGV; Luque *et al.*, 2001), *Adoxophyes orana GV* (AdorGV; Wormleaton *et al.*, 2003) and *Cryptophlebia leucotreta GV* (CrleGV; Lange & Jehle, 2003). Based on phylogeny of the 30 core genes and the presence (group I) or absence (group II) of the *gp67*/*gp64* fusion protein gene (Pearson *et al.*, 2000), 14 published NPV genomes include six group I NPVs; *Autographa californica MNPV* (AcMNPV; Ayres *et al.*, 1994), *Bombyx mori NPV* (BmNPV; Gomi *et al.*, 1999), *Choristoneura fumiferana* defective NPV, *Epiphyas postvittana NPV* (EppoNPV; Hyink *et al.*, 2002), *Orgyia pseudotsugata MNPV* (OpMNPV; Ahrens *et al.*, 1997) and *Rachiplusia ou NPV* (RoMNPV; Harrison & Bonning, 2003), and eight group II NPVs, *Helicoverpa armigera NPV* (HearNPV G4; Chen *et al.*, 2001), *Helicoverpa zea single NPV* (HzSNPV; Chen *et al.*, 2002), *Lymantria dispar MNPV* (LdMNPV; Kuzio *et al.*, 1999), *Mamestra configurata NPV* A (MacoNPV A; Q. Li *et al.*, 2002), *Mamestra configurata NPV* B (MacoNPV B; L. Li *et al.*, 2002), *Spodoptera exigua MNPV*

The GenBank/EMBL/DDBJ accession number of the complete genome sequence of CfMNPV reported in this paper is AF512031.

(SeMNPV; IJkel *et al.*, 1999), Spodoptera litura NPV (SpltNPV; Pang *et al.*, 2001) and *Adoxophyes honmai NPV* (AdhoNPV; Nakai *et al.*, 2003).

Choristoneura fumiferana MNPV (CfMNPV) is a group I multiple encapsidated NPV, infectious to the eastern spruce budworm, *Choristoneura fumiferana*, which historically has been one of the most destructive forest insect pest species in North America destroying up to 35 million hectares of forest per year. Arif *et al.* (1984) constructed a physical map for the CfMNPV genome estimated at 124·4–126·4 kb in size. The sequences of CfMNPV genes *p143*, *lef-3*, *iap-2*, *vlf-1* (Chen *et al.*, 2004), *ie-1*, *ie-2*, *pe38* (Carstens *et al.*, 2002), *pkip*, *p47*, *lef-12*, *gta* (Lapointe *et al.*, 2000), *p48*, *p82* (Li *et al.*, 1997, 1999), *slp* (Liu & Carstens, 1996), *DNApol* (Liu & Carstens, 1995), *cathepsin*, *gp67*/*gp64* (Hill *et al.*, 1993, 1995) and *p10* (Wilson *et al.*, 1995) have been published and clustered with those of OpMNPV by phylogenetic analysis. This paper describes the complete CfMNPV genome sequence and organization, compares it to that of other published baculovirus genomes and places it in the context of baculovirus phylogeny.

METHODS

Virus preparation, construction of genomic DNA libraries and DNA sequencing. Viral DNA was isolated from plaque purified viral stocks (Arif *et al.*, 1984; Liu & Carstens, 1993) and cloned as *Hin*dIII, *Eco*RI or *Bam*HI fragments into pUC18 or 19 (*Hin*dIII) or pBR322 (*Eco*RI and *Bam*HI). Sequence was obtained from both strands using plasmid specific primers and 'primer walking'. PCR fragments of virion DNA covering the flanking regions of restriction enzyme sites used for cloning were also sequenced. Sequencing was accomplished by dideoxynucleotide chain termination using the BigDye Terminator system (Applied BioScience), and the Applied Biosystems 377 Prism sequencer. Sequence represents at least triple redundancy. Sequencing was done at the Molecular Super Centre at the University of Guelph, Guelph, ON, Canada. GenBank submission sequences U26675, AF067799 and AF127271 were also utilized.

Sequence analysis. Sequence data were compiled into contigs using GeneRunner (www.generunner.com). Open reading frames (ORFs) were identified using ORF finder (http://www.ncbi.nlm.nih.gov/gorf). The criterion for defining an ORF was a size of at least 50 aa with minimal overlap. All BLAST searches were done through the National Centre for Biotechnology Information (NCBI) website using BLAST 2.2.3. Multiple alignments and percentage identities were generated by using the alignX package from Vector NTI (Invitrogen). Genome parity plots were generated using GenBank data and as described previously (Hu *et al.*, 1998).

RESULTS AND DISCUSSION

Nucleotide sequence and analysis of the CfMNPV genome

The entire CfMNPV dsDNA genome was sequenced and assembled into a contiguous sequence of 129 593 bp with a G+C content of 50·1 mol%. This size was slightly larger than the estimate of 124·4–126·4 kb based on restriction fragment sizes (Arif *et al.*, 1984). The first nucleotide of the sequence was designated the last adenine of the *polyhedrin* stop codon with *polyhedrin* in the reverse orientation. The size of sequenced baculovirus genomes ranges from 82 (NeleNPV; Lauzon *et al.*, 2004) to 178 kb (XecnGV; Hayakawa *et al.*, 1999) with G+C contents ranging from 32·4 (CrleGV; Lange & Jehle, 2003) to 57·7 mol% (LdMNPV; Kuzio *et al.*, 1999).

A total of 146 methionine-initiated ORFs, with no or minimal overlap and encoding putative proteins of 50 aa or more were identified (Fig. 1, Table 1). Exceptions regarding overlap were made for CfORFs 2, 40, 51 and 76 based on the identity and length of their homologues in other baculoviruses. The CfMNPV ORFs demonstrated no preference to orientation (49 % forward and 51 % reverse) or clustering based on function or expression, which was consistent with other baculovirus genomes. Of the 146 identified ORFs, 139 had homologues in at least one other baculovirus and seven were unique to CfMNPV. Five regions that resembled baculovirus homologous regions (*hr*s) were also identified. *hr*s have been implicated as origins of DNA replication (Ahrens & Rohrmann, 1995; Kool *et al.*, 1995; Lee & Krell, 1994) and transcriptional enhancers (Theilmann & Stewart, 1992). ORFs accounted for 117 551 bp, *hr*s for 2762 bp and the remaining 9280 bp represented intergenic or unidentified regions. Most of the intergenic regions ranged from 0 to 200 bp in length, but were as long as 762 bp (Cf97/Cf98). However, some contiguous ORFs also overlapped, in 29 cases by as little as 2–79 bp, but longer overlaps of 119 (Cf76/Cf77), 152 (Cf51/Cf52), 199 (Cf2/Cf3) and 416 bp (Cf40/Cf41) were also observed (Table 1).

Comparison of CfMNPV gene content with other baculoviruses

CfMNPV shared ORFs 131, 128, 133, 125 and 122 with the other group I NPVs, OpMNPV, EppoNPV, CfDEFNPV, AcMNPV, BmNPV and RoMNPV, respectively. A total of 117 ORFs were conserved among all six group I NPVs. Herniou *et al.* (2003) and Hyink *et al.* (2002) identified 13 ORFs exclusive to group I NPVs including *lef-7*. However, due to the presence of a *lef-7* homologue in the group II NPV, MacoNPV A (MacoORF16) (Q. Li *et al.*, 2002), only 12 ORFs are now exclusive to and present in all group I NPVs, including the *gp64*/*gp67* fusion protein gene required for cell-to-cell transmission of group I NPVs (Monsma *et al.*, 1996), transcriptional activator genes such as *ie-2* (Lu & Miller, 1995) and *iap-1*, encoding an inhibitor of apoptosis (Maguire *et al.*, 2000).

Compared to the group II NPVs, CfMNPV shared the highest number of ORFs with MacoNPV A (94 ORFs) and the least with LdMNPV and SpltMNPV (84 ORFs). A total of 74 baculovirus ORFs were present in all lepidopteran NPVs, including CfMNPV. Some of these ORFs were also present in one or more GVs. Of the 74 ORFs, eight were found exclusively in lepidopteran NPVs, [ac17, ac21 (*actin rearrangement inducing factor*, *arif-1*), ac34, ac55, ac57,

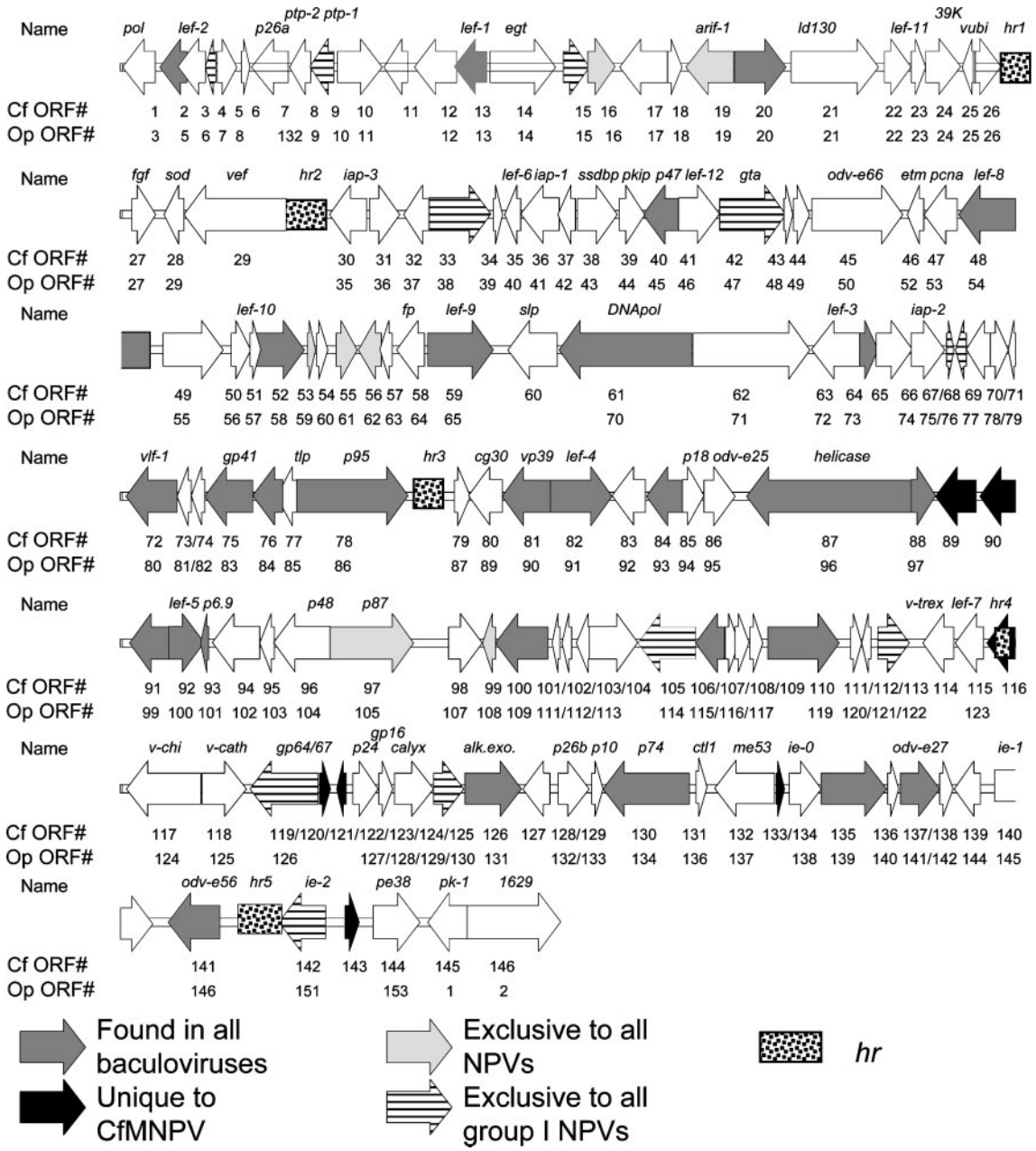

Fig. 1. CfMNPV ORF map. ORFs are designated by arrows, indicating direction of transcription. Homologous regions (*hr*s) are defined by stippled boxes.

ac59, ac104, ac108] of which only *arif-1* (Cf19, ac21) was studied in detail. CfMNPV shared 69, 65 and 68 ORFs with the granuloviruses CpGV, PlxyGV and XecnGV, respectively.

CfMNPV had homologues of two ORFs identified as unique to EppoNPV. Multiple alignment of Cf11 to EppoNPV ORF9 (15·2 % amino acid identity) revealed areas of high conservation including a strongly conserved C terminus (71·2 % over the last 12 aa). Cf103 was 34·5 % identical at the amino acid level to Eppo98. Both Cf103 and Eppo98 had a 23 residue C-4 zinc finger motif (aa 79–101 in CfMNPV and aa 75–97 in EppoNPV). The percentage identity within these regions (78·3 %) was much higher than the overall amino acid identity (34·5 %), indicating that this region may be functionally important.

Genes involved in DNA replication

Six baculovirus genes, *lef-1*, *lef-2*, *lef-3*, *DNApol*, *p143* (*hel*) and *ie-1*, are essential in transient assays for DNA replication in AcMNPV and OpMNPV and exist in all lepidopteran baculoviruses including CfMNPV (Kool *et al.*, 1995; Lu & Miller, 1995; Ahrens & Rohrmann, 1995). Only four are

Table 1. ORFs identified in CfMNPV

CfMNPV ORF number is given in the first column. The right and left boundaries are given by nucleotide number in the 'position' column and the direction of transcription is given by the <(−ve strand) and >(+ve strand) symbols. The intergenic distance between the ORFs is given in bp, a negative number indicates overlap. The ORF numbers and amino acid identities are given for OpMNPV, EppoNPV, AcMNPV, BmNPV, MacoNPV A and XecnGV. The unique ORFs are underlined and in bold. Note in cases where there are two or more homologues in one virus (i.e. MacoNPV A also has two copies of *p26*) the ORF # showing the highest % identity is given in the table.

Cf ORF#	Name	Position	Intergenic distance (bp)	Size (amino acid)	ORF# (amino acid identity %)					
					OpMNPV	EppoNPV	AcMNPV	BmNPV	MacoNPV A	XecnGV
1	*pol*	1<735	0	245	3 (98·0)	1 (97·1)	8 (87·8)	1 (89·8)	1 (84·1)	1 (52·8)
2		838<1449	102	204	5 (49·3)	3 (45·3)	–	–	–	–
3	*lef-2*	1251<1859	−199	203	6 (80·9)	4 (68·4)	6 (53·8)	135 (53·8)	14 (36·8)	35 (13·6)
4		1862<2092	2	77	7 (81·8)	5 (36·1)	5 (28·4)	134 (26·6)	–	–
5		2137>2568	44	144	8 (66·4)	6 (54·1)	4 (55·0)	133 (55·0)	–	–
6	*bro-a*	2669>2860	0	64	–	–	–	–	–	–
7	*p26a*	2896<3699	35	268	132 (26·2)	119 (27·2)	136 (31·3)	113 (31·0)	158 (48·3)	–
8	*ptp-2*	3738<4220	38	161	9 (71·9)	–	–	–	38 (22·2)	–
9	*ptp-1*	4198<4731	−24	178	10 (64·1)	7 (65·0)	1 (56·5)	130 (56·5)	–	–
10		4800>5819	68	340	11 (64·6)	8 (48·7)	11 (38·8)	9 (38·9)	4 (39·4)	–
11		5854<6387	34	178	–	9 (15·2)	–	–	–	–
12		6527<7483	139	319	12 (73·4)	10 (60·1)	13 (44·0)	5 (42·3)	36 (20·3)	–
13	*lef-1*	7426<8166	−58	247	13 (73·2)	11 (59·3)	14 (50·4)	6 (50·4)	35 (30·5)	82 (28·0)
14	*egt*	8235>9710	68	492	14 (79·6)	12 (71·1)	15 (59·8)	7 (59·2)	39 (43·0)	–
15		9864>10457	153	198	15 (70·1)	13 (48·8)	16 (28·4)	8 (27·5)	–	–
16		10426>11049	−32	208	16 (72·0)	14 (60·8)	17 (47·1)	9 (48·1)	41 (22·7)	–
17		11139<12197	89	353	17 (73·0)	15 (63·4)	18 (50·0)	10 (47·5)	56 (19·2)	–
18		12210>12524	12	105	18 (70·2)	16 (58·5)	19 (42·6)	11 (37·3)	–	–
19	*arif-1*	12631<13683	106	351	19 (51·4)	17 (42·8)	21 (29·7)	12 (27·7)	47 (15·5)	–
20		13708>14859	24	384	20 (89·8)	18 (84·3)	22 (79·1)	13 (79·9)	48 (49·9)	45 (47·3)
21	*copia*	14970>16934	110	655	21 (69·4)	19 (47·5)	23 (34·8)	14 (35·2)	9 (15·9)	27 (13·0)
22		17060>17665	125	202	22 (85·0)	20 (79·7)	38 (69·9)	29 (69·5)	148 (51·1)	79 (34·5)
23	*lef-11*	17673>17996	7	108	23 (69·6)	21 (57·9)	37 (49·1)	28 (47·3)	149 (27·3)	56 (21·2)
24	*39K*	17990>18772	−7	261	24 (79·4)	22 (70·7)	36 (48·0)	27 (49·1)	150 (25·4)	55 (17·8)
25	*v-ubi*	18802<19089	29	95	25 (67·9)	23 (75·5)	35 (71·3)	26 (71·3)	152 (54·3)	52 (63·1)
26		19058>19684	−32	209	26 (78·9)	24 (63·5)	34 (51·6)	25 (52·1)	153 (34·9)	–
	hr1	19697-20116	12							
27	*fgf*	20212>20751	95	180	27 (62·9)	25 (60·2)	32 (39·6)	24 (38·3)	51 (12·8)	144 (12·9)
28	*sod*	20931<21389	179	153	29 (82·2)	–	31 (76·3)	23 (73·7)	66 (70·4)	68 (52·6)
29	*vef*	21396<23672	6	759	–	–	–	–	89 (16·8)	150 (15·7)
	hr2	23798–24453	125							
30	*iap-3*	24635<25480	181	282	35 (57·8)	26 (54·6)	–	–	139 (37·6)	
31		25558>26253	52	229	36 (68·0)	27 (33·3)	–	–	–	–
32		26303<26866	99	188	37 (45·5)	29 (43·1)	–	–	–	–
33		26865>28238	−2	458	38 (77·6)	30 (62·9)	30 (51·2)	21 (49·6)	–	–
34		28293>28499	54	69	39 (57·3)	31 (57·4)	29 (50·7)	20 (47·9)	157 (25·9)	–
35	*lef-6*	28551<28925	51	125	40 (50·3)	32 (42·5)	28 (21·8)	19 (21·2)	156 (25·0)	88 (12·1)
36	*iap-1*	28925<29755	0	277	41 (83·3)	33 (69·0)	27 (55·9)	18 (53·8)	–	–
37		29752<30129	−4	126	42 (79·5)	34 (67·7)	26 (58·9)	17 (55·8)	154 (20·5)	–
38	*ssdbp*	30184>31086	54	301	43 (88·3)	35 (69·3)	25 (41·5)	16 (40·1)	155 (20·7)	89 (18·2)
39	*pkip*	31096>31596	9	167	44 (79·5)	36 (57·0)	24 (45·9)	15 (44·1)	–	–
40	*p47*	31650<32849	53	400	45 (83·0)	37 (78·7)	40 (65·8)	31 (65·6)	145 (50·0)	78 (40·7)
41	*lef-12*	32434>33348	−416	305	46 (44·7)	38 (36·9)	41 (24·5)	32 (25·9)	–	–
42	*gta*	33357>34850	8	498	47 (86·7)	39 (70·5)	42 (59·6)	33 (59·6)	–	–
43		34853>35032	2	60	48 (65·1)	40 (45·3)	43 (40·3)	34 (46·2)	–	–
44		35017>35382	−16	122	49 (63·9)	41 (73·6)	44 (37·9)	35 (38·6)	–	–
45	*odv-e66*	35434>37470	51	679	50 (74·5)	42 (76·5)	46 (67·4)	37 (68·3)	78 (35·2)	149 (36·5)

Table 1. cont.

Cf ORF#	Name	Position	Intergenic distance (bp)	Size (amino acid)	ORF# (amino acid identity %)					
					OpMNPV	EppoNPV	AcMNPV	BmNPV	MacoNPV A	XecnGV
46	*etm*	375552<37929	81	126	52 (68·0)	44 (50·8)	48 (36·4)	–	–	–
47	*pcna*	37943<38677	13	245	53 (62·2)	–	49 (33·5)	–	–	–
48	*lef-8*	38708<41332	30	875	54 (86·7)	45 (80·3)	50 (69·0)	39 (68·9)	141 (56·3)	148 (46·5)
49		41365>42300	32	312	55 (53·7)	46 (44·7)	51 (27·2)	40 (27·2)	–	–
50		42429>42869	128	147	56 (83·6)	47 (77·4)	53 (54·8)	42 (56·2)	137 (39·8)	171 (23·2)
51	*lef-10*	42838>43080	−32	81	57 (71·3)	48 (60·5)	53a (42·7)	42a (43·9)	–	–
52	*vp1054*	42929>44065	−152	379	58 (83·7)	49 (49·5)	54 (49·5)	43 (49·5)	133 (35·2)	175 (25·3)
53		44127>44321	61	65	59 (66·2)	50 (62·9)	55 (47·3)	44 (42·3)	–	–
54		44334>44582	12	83	60 (80·5)	51 (80·5)	56 (48·8)	45 (45·2)	–	–
55		44768>45250	185	161	61 (76·1)	52 (45·7)	57 (50·0)	46 (48·8)	130 (34·6)	–
56		45274<45780	23	169	62 (62·7)	53 (47·0)	59 (22·6)	47 (35·7)	129 (28·7)	–
57		45752<46021	−35	90	63 (80·0)	54 (58·4)	60 (51·7)	48 (49·4)	128 (34·7)	102 (21·8)
58	*fp*	46134<46760	112	209	64 (87·5)	55 (78·8)	61 (69·3)	49 (70·9)	125 (51·9)	140 (21·4)
59	*lef-9*	46824>48296	63	491	65 (91·6)	56 (85·3)	62 (72·3)	50 (76·4)	124 (62·7)	139 (52·1)
60	*slp*	48619<49719	322	367	69 (64·2)	57 (83·3)	64 (60·4)	52 (58·2)	37 (40·3)	107 (27·8)
61	*DNApol*	49766<52738	46	991	70 (82·5)	58 (73·4)	65 (60·9)	53 (60·1)	115 (40·1)	132 (28·4)
62		52748>55384	9	879	71 (55·1)	59 (34·2)	66 (26·5)	54 (25·8)	114 (17·4)	–
63	*lef-3*	55381<56502	−4	374	72 (75·7)	60 (60·1)	67 (38·5)	55 (38·8)	113 (20·5)	134 (12·8)
64		56501>56896	−2	132	73 (84·0)	61 (67·9)	68 (42·7)	56 (61·2)	112 (38·2)	135 (22·4)
65		56868>57668	−29	267	–	62 (76·5)	69 (55·6)	57 (55·6)	111 (37·8)	–
66	*iap-2*	57649>58404	−20	252	74 (71·7)	63 (65·5)	71 (56·6)	58 (57·5)	110 (30·0)	–
67		58437>58607	32	57	75 (66·1)	64 (46·7)	72 (52·5)	58a (47·5)	–	–
68		58638<58886	30	83	76 (37·9)	65 (34·5)	73 (20·2)	59 (20·2)	–	–
69		58883<59404	−4	174	77 (72·8)	66 (52·2)	74 (25·0)	60 (24·7)	–	–
70		59417<59809	12	131	78 (84·6)	67 (51·5)	75 (42·1)	61 (42·9)	116 (24·2)	–
71		59814<60068	4	85	79 (95·2)	68 (91·7)	76 (83·3)	62 (80·0)	117 (40·0)	125 (28·2)
72	*vlf-1*	60080<61204	11	375	80 (93·9)	69 (89·1)	77 (78·4)	63 (77·3)	106 (58·9)	123 (28·9)
73		61206<61526	1	107	81 (74·5)	70 (73·8)	78 (60·6)	64 (59·1)	105 (20·8)	122 (25·0)
74		61523<61837	−4	105	82 (91·3)	71 (76·9)	79 (65·1)	65 (65·1)	17 (40·0)	75 (28·3)
75	*gp41*	61841<62926	3	362	83 (85·3)	72 (82·8)	80 (61·9)	66 (60·0)	104 (41·8)	121 (23·5)
76		62919<63575	−8	219	84 (90·8)	73 (83·0)	81 (62·2)	67 (60·7)	103 (29·6)	120 (37·9)
77	*tlp*	63457<63924	−119	156	85 (76·8)	74 (52·9)	82 (28·6)	68 (27·3)	102 (16·9)	119 (17·9)
78	*vp91*	63893>66394	−32	834	86 (79·6)	75 (72·1)	83 (61·6)	69 (58·7)	101 (41·0)	118 (15·2)
	hr3	66508–66958	97							
79	*capsid*	67424>67786	481	121	88 (73·3)	–	87 (39·3)	70 (39·7)	–	–
80	*cg30*	67746<68495	−41	250	89 (76·7)	76 (52·6)	88 (43·1)	71 (41·6)	100 (14·6)	–
81	*vp39*	68501<69577	5	359	90 (81·7)	77 (66·7)	89 (59·3)	72 (58·1)	99 (39·7)	111 (23·4)
82	*lef-4*	69588>70961	10	458	91 (81·4)	78 (65·4)	90 (51·5)	73 (50·3)	98 (38·5)	110 (29·2)
83		70948<71703	−14	252	92 (61·2)	79 (63·7)	91 (51·6)	74 (33·1)	–	–
84		71719<72531	15	271	93 (80·9)	80 (79·3)	92 (79·7)	75 (79·3)	96 (46·9)	101 (33·2)
85	*p18*	72530>73009	−2	160	94 (94·3)	81 (83·0)	93 (71·4)	76 (70·8)	95 (43·6)	100 (21·9)
86	*odv-e25*	73014>73703	5	230	95 (89·5)	82 (86·9)	94 (63·9)	77 (60·0)	94 (39·6)	99 (33·3)
87	*hel*	73969<77655	265	1229	96 (84·7)	83 (71·5)	95 (56·9)	78 (55·9)	93 (35·3)	98 (21·0)
88		77645>78202	−11	186	97 (75·1)	84 (76·2)	96 (63·4)	79 (62·9)	92 (39·4)	97 (25·8)
89		**78199<79104**	**−4**	**302**	–	–	–	–	–	–
90		**79170<80024**	**65**	**285**	–	–	–	–	–	–
91		80149<81078	124	310	99 (81·5)	85 (75·1)	98 (56·9)	82 (57·5)	88 (41·0)	96 (35·2)
92	*lef-5*	81025>81810	−54	262	100 (79·8)	86 (71·2)	99 (58·1)	83 (57·0)	87 (44·4)	95 (30·5)
93	*p6·9*	81807<81959	−4	51	101 (90·2)	87 (83·0)	100 (69·1)	84 (58·5)	86 (44·7)	94 (39·3)
94		82001<83059	41	353	102 (82·5)	88 (74·5)	101 (59·8)	85 (59·4)	85 (37·3)	93 (18·4)
95		83069<83404	9	112	103 (84·8)	89 (81·3)	102 (42·6)	86 (39·8)	84 (25·2)	92 (23·3)
96	*p48*	83382<84617	−23	412	104 (87·1)	90 (76·9)	103 (56·1)	87 (55·8)	83 (39·2)	91 (32·0)
97	*p87*	84641>86515	23	625	105 (55·1)	91 (37·3)	104 (30·3)	88 (29·3)	82 (15·1)	–

Table 1. cont.

Cf ORF#	Name	Position	Intergenic distance (bp)	Size (amino acid)	ORF# (amino acid identity %)					
					OpMNPV	EppoNPV	AcMNPV	BmNPV	MacoNPV A	XecnGV
98		87278>88045	762	256	107 (75·0)	93 (66·7)	106/107 (15·7)	90 (58·9)	71 (41·8)	50 (33·0)
99		88052<88354	6	101	108 (81·5)	94 (73·3)	108 (49·5)	91 (50·5)	79 (18·2)	–
100		88357<89529	2	391	109 (89·7)	95 (84·0)	109 (66·8)	92 (65·6)	80 (40·3)	53 (29·7)
101		89609<89779	79	57	111 (89·3)	96 (70·2)	110 (69·6)	92a (67·8)	81 (32·8)	51 (22·4)
102		89830<90042	50	71	112 (75·0)	97 (67·1)	111 (60·0)	93 (60·0)	–	160 (28·0)
103		90141<90506	98	122	–	98 (34·5)	–	–	–	–
104		90442>91527	−62	362	113 (76·1)	99 (64·1)	–	–	–	–
105		91543<92817	15	425	114 (71·6)	100 (52·8)	114 (36·3)	94 (37·0)	–	–
106		92828<93460	10	211	115 (75·2)	102 (65·2)	115 (59·5)	95 (60·5)	68 (41·3)	32 (34·0)
107		93490>93780	29	97	116	103 (24·1)	–	–	–	–
108		93708>94013	−73	102	116 (42·0)	103 (42·0)	–	–	–	–
109		94043>94336	29	98	117 (69·1)	104 (39·2)	117 (30·8)	96 (30·8)	62 (17·6)	–
110		94445>96037	108	531	119 (83·4)	106 (75·1)	119 (74·6)	97 (69·9)	49 (47·1)	84 (29·6)
111		96294>96542	256	83	120 (75·6)	107 (57·0)	120 (51·2)	98 (52·4)	–	–
112		96546<96746	3	67	121 (51·4)	–	122 (30·4)	99 (29·9)	–	–
113		96892>97605	145	238	122 (62·1)	108 (53·1)	124 (38·1)	101 (38·1)	–	–
114	*v-trex*	97903<98601	297	233	–	–	–	–	–	–
115	*lef-7*	98641<99282	39	214	123 (56·8)	109 (24·2)	125 (26·9)	102 (26·5)	16 (16·4)	–
116		**99333<100112**	**50**	**269**	–	–	–	–	–	–
	hr4	99565–100090	–	–	–	–	–	–	–	–
117	*v-chi*	100081<101739	−32	553	124 (86·2)	110 (79·3)	126 (80·8)	103 (79·9)	22 (64·2)	103 (54·4)
118	*v-cath*	101783>102757	43	325	125 (82·4)	111 (74·4)	127 (79·6)	104 (78·4)	33 (53·1)	58 (57·2)
119	*gp64*	102843<104372	85	510	126 (88·4)	112 (78·8)	128 (79·1)	105 (76·2)	–	–
120		**104399>104650**	**26**	**84**	–	–	–	–	–	–
121		**104790<104984**	**139**	**65**	–	–	–	–	–	–
122	*p24*	105140>105712	156	191	127 (84·4)	114 (75·3)	129 (64·1)	106 (61·0)	12 (33·5)	80 (17·6)
123	*gp16*	105725>106033	12	103	128 (83·5)	115 (73·5)	130 (65·1)	107 (65·1)	11 (26·5)	–
124	*pep*	106081>106950	47	290	129 (73·8)	116 (76·4)	131 (44·1)	108 (57·4)	60 (31·2)	–
125		106952>107632	1	227	130 (54·6)	117 (38·0)	132 (23·7)	109 (25·4)	–	–
126	*alkexo*	107647>108918	14	424	131 (79·7)	118 (67·3)	133 (51·6)	110 (51·2)	54 (35·8)	145 (27·1)
127		108952<109560	33	203	–	–	–	–	–	–
128	*p26b*	109748>110449	184	234	132 (64·4)	119 (53·5)	136 (51·7)	113 (50·4)	158 (27·6)	–
129	*p10*	110500>110745	50	82	133 (43·0)	120 (38·8)	137 (43·6)	114 (39·4)	159 (30·0)	5 (29·4)
130	*p74*	110746<112683	0	646	134 (89·8)	121 (85·7)	138 (78·0)	115 (77·9)	160 (49·0)	77 (31·7)
131	*ctl-1*	113034>113195	350	54	136 (66·0)	–	3 (81·1)	–	107 (45·3)	127 (41·5)
132	*me53*	113240<114586	44	449	137 (77·0)	122 (56·6)	139 (36·3)	116 (34·8)	7 (18·1)	180 (13·8)
133		**114617>114799**	**30**	**61**	–	–	–	–	–	–
134	*ie-0*	114913>115647	113	245	138 (73·1)	123 (65·2)	141 (50·0)	117 (49·6)	168 (23·6)	–
135		115660>117099	12	480	139 (91·5)	124 (85·0)	142 (71·6)	118 (71·4)	167 (46·5)	13 (31·1)
136	*odv-e18*	117121>117378	51	86	140 (91·8)	125 (86·2)	143 (42·2)	119 (54·5)	166 (34·8)	12 (30·0)
137	*odv-e27*	117407>118297	28	297	141 (78·5)	126 (84·6)	144 (69·4)	120 (69·7)	165 (44·9)	112 (25·0)
138		118301>118588	3	96	142 (92·6)	127 (82·1)	145 (56·8)	121 (68·4)	164 (46·3)	87 (37·3)
139		118618<119211	29	198	144 (69·5)	128 (65·7)	146 (50·2)	122 (49·8)	163 (27·4)	10 (21·1)
140	*ie-1*	119264>120946	52	561	145 (72·4)	129 (66·0)	147 (44·0)	123 (44·0)	162 (24·0)	9 (10·8)
141	*odv-e56*	121019<122158	72	380	146 (84·4)	130 (81·1)	148 (67·7)	124 (65·8)	6 (48·2)	15 (34·5)
	hr5	122228–122936	53							
142	*ie-2*	123494<124540	573	349	151 (36·3)	131 (34·0)	151 (27·1)	127 (25·1)	–	–
143		**124982>125284**	**441**	**101**	–	–	–	–	–	–
144	*pe38*	125600>126634	315	345	152 (37·6)	133 (33·8)	153 (22·6)	128 (22·6)	–	–
145	*pk-1*	126825<127700	190	292	1 (79·1)	135 (70·8)	10 (59·5)	3 (60·5)	3 (31·3)	3 (24·2)
146	*cap*	127699>129593	−2	631	2 (32·4)	136 (34·6)	9 (27·2)	2 (27·0)	2 (15·1)	–
Average					**73·6**	**62·9**	**51·5**	**51·5**	**35·3**	**28·5**

conserved among all sequenced baculovirus genomes as CuniNPV lacks recognizable *ie-1* and *lef-3* homologues (Afonso *et al.*, 2001). CuniNPV infects a dipteran, rather than a lepidopteran host, which might account for these differences. *DNApol* was the most highly conserved gene involved in DNA replication with a mean identity of 43·6 % at the amino acid level when the CfMNPV DNApol was compared with those from all sequenced baculovirus genomes. This high degree of conservation of DNApol may reflect the need to conserve its functional domains (e.g. nucleotide binding/5′-3′ polymerization) (Liu & Carstens, 1995). The least conserved gene product of this group was lef-3 (identity of 28 %), a protein with single-stranded DNA binding abilities (Chen *et al.*, 2004; Hang *et al.*, 1995) and thought to be a chaperone for the transport of P143 helicase to the nucleus (Wu & Carstens, 1998; Chen *et al.*, 2004). Homologues of an additional single-stranded DNA-binding protein (*ssdbp*/ac25) (Mikhailov *et al.*, 1998) and immediate-early gene *me53*/ac139, both of which have been implicated in DNA replication, were also found in CfMNPV (Cf38 and Cf132). Homologues of the non-essential DNA replication stimulatory genes *ie-2*, *lef-7* and *pe38* (Kool *et al.*, 1994) were found in CfMNPV (Carstens *et al.*, 2002). Several baculoviruses, including OpMNPV, MacoNPV A, MacoNPV B, LdMNPV, SeMNPV, SpltMNPV and CpGV encode ribonucleotide reductase subunits and/or a dUTPase, which are involved in nucleotide metabolism. These genes may allow the viruses to replicate in non-dividing cells in which the nucleotide biosynthesis pathways have been shut-off (Ahrens *et al.*, 1997). However, no homologues to *rr1*, *rr2* or *dutpase* or any other gene involved in nucleotide metabolism were found in CfMNPV.

Genes regulating transcription

Baculovirus gene transcription occurs in a temporal cascade for the immediate-early, delayed-early, late and very late genes. Immediate-early and delayed-early gene expression occurs prior to DNA replication and utilizes host RNA polymerase II while late and very late gene expression occurs after initiation of DNA replication and is driven by a viral encoded RNA polymerase (Miller, 1997). In AcMNPV the viral RNA polymerase comprises four subunits encoded by *lef-4*, *lef-8*, *lef-9* and *p47* (Guarino *et al.*, 1998) and these are present in all fully sequenced baculovirus genomes, including that of CfMNPV. These were highly homologous with mean amino acid identities of 40·4–62·3 % when compared with those from CfMNPV. The most highly conserved transcriptional protein was lef-9 (mean identity of 62·3 %), which is the subunit containing the polymerization domain. Homologues of both *lef-5* and *vlf-1*, which are present in all baculovirus genomes sequenced to date, were also found in CfMNPV. Although the function of *lef-5* is unclear, *vlf-1* is essential for the 'burst' in very late gene expression seen for *p10* and *polyhedrin* (Yang & Miller, 1998, 1999).

In addition to the six transcriptional genes described above, CfMNPV encodes homologues of *39K*/*pp31*, *lef-6* and *lef-11*, which are present in all sequenced lepidopteran baculoviruses (Herniou *et al.*, 2003). The early gene transactivators me53, ie-0 and ie-1, also present in CfMNPV, were generally less conserved with mean identities of 30, 37 and 35 %, respectively. Since the corresponding proteins interact with host factors, such as host RNA polymerase, rather than viral factors, they may have evolved to be more host specific, resulting in a greater degree of variation among these genes from different baculoviruses (IJkel *et al.*, 1999).

Inhibitors of apoptosis

Like both OpMNPV and EppoNPV, CfMNPV encoded homologues of *iap-1*, *iap-2* and *iap-3*, but lacked the *p35* caspase inhibitor homologue found in AcMNPV, RoMNPV and BmNPV. This absence is not surprising as other members of the *iap* family are functionally equivalent to *p35* from AcMNPV (Crook *et al.*, 1993). The CfMNPV *iap* gene products had high sequence identity to IAPs from OpMNPV, Hyphantria cunea NPV (HycuNPV) and EppoNPV. CfMNPV IAP-1, -2 and -3 possess C-terminal zinc RING-finger domains common to IAPs (Crook *et al.*, 1993). CfMNPV *iap-3* also had two tandem copies of the baculovirus *iap* repeat (BIR) sequence and *iap-1* had one BIR sequence (Birnbaum *et al.*, 1994). CfMNPV lacked *iap-4*, which is present in both OpMNPV and EppoNPV, the viruses that are most closely related to CfMNPV.

Structural genes

Herniou *et al.* (2003) identified nine structural genes common to 13 sequenced baculovirus genomes, all of which were found in CfMNPV (*ld130*, *gp41*, *odv-e27*, *odv-e56*, *p6·9*, *p74*, *vp91*, *vp39*, *vp1054*). Furthermore, CfMNPV has six additional structural protein genes found in all lepidopteran baculoviruses (*fp25K*, *odv-e18*, *odv-e25*, *odv-e66*, *pk-1*, *polyhedrin*/*granulin*). Also found in CfMNPV were several NPV-specific genes including *polyhedron envelope protein* (*pep*) (Cf124), *vp*80/87 (Cf97) and *slp* (Cf60) and the group I specific *gp67*.

Auxiliary genes

Baculovirus genomes typically encode auxiliary genes that are non-essential for replication but otherwise provide a selective advantage to the virus (Miller, 1997). Several of these auxiliary genes have homologues in CfMNPV, including *protein tyrosine phosphatase-1* (*ptp-1*), *ptp-2*, *ecdysteroid UDP glucosyltransferase* (*egt*), *arif*, *ubiquitin*, *fibroblast growth factor* (*fgf*), *superoxide dismutase* (*sod*), *viral enhancing factor* (*vef*), *proliferating cell nuclear antigen* (*pcna*), *viral chitinase* (*v-chi*), *viral cathepsin* (*v-cath*), *alkaline exonuclease* and *conotoxin-like peptide*. The only auxiliary gene conserved among all baculoviruses is *alkaline exonuclease*, which interacts with lef-3 and can degrade both single- and double-stranded DNA in a 5′-3′ direction and possesses both endo- and exonuclease activities (Mikhailov *et al.*, 2003). The *alkaline exonuclease* may be involved in the processing of replicative intermediates

Table 2. Summary and description of CfMNPV unique ORFs

ORF name	Size (amino acid)	Molecular mass (kDa)	pI	Putative promoters
Cf89	301	36·89	9·16	CAGT: −188
Cf90	284	34·04	8·94	CAGT: −21 CATT: −122, −132, −158
Cf116*	259	31·29	12·06	CAGT: −110 CATT: −185
Cf120	83	9·40	10·56	CAGT: −39 CATT: −101
Cf121	64	7·14	4·95	CAGT: −95
Cf133	60	6·92	8·90	ND
Cf143	100	10·98	9·74	TTAAG: −25

*Cf116 overlaps with *hr4*.
ND, Not detected.

(Li & Rohrmann, 2000) and/or repair and recombination of viral genomes (Mikhailov *et al.*, 2003) during infection. *fgf* and *ubiquitin* are also conserved among all lepidopteran baculoviruses, however their roles in baculovirus infection are unknown.

Unique ORFs found in CfMNPV

Seven CfMNPV ORFs had no recognizable baculovirus or other GenBank homologues (Table 2). Of these ORFs, Cf143 had an upstream late promoter motif (TTAAG) 25 nt upstream of the start codon. Six ORFs had conserved baculovirus early promoter motifs (CAKT) 200 bases or less upstream of their ATGs, while no baculovirus promoter motif was detected upstream of Cf133.

ORFs demonstrating significant protein motifs

BLAST analysis of Cf114 revealed the presence of exonuclease signature patterns. Multiple alignment of Cf114 with members of the TREX family of 3′-5′ exonucleases identified three domains similar to the conserved motifs for the defined ExoI [DxE(S/T/C)], ExoII [Nx_{2-3}(F/Y)D] and ExoIII ($HxAx_2D$) functional domains (Yang *et al.*, 2004). Although Cf114 is the first example of a baculovirus *v-trex* ORF homologue reported in a complete baculovirus genome one with 67·5 % identity to Cf114 was also reported in a partial genome fragment from AgMNPV DNA (Slack *et al.*, 2004) and a homologue exists in CfDEFNPV. The DNA polymerase from Spodoptera litoralis NPV (SpliNPV; Huang & Levin, 2001), also demonstrated intrinsic 3′-5′ exonuclease activity. Typically 3′-5′ exonucleases act as 'proofreading' enzymes during or after DNA replication. Perhaps the CfMNPV DNA polymerase lacks intrinsic 3′-5′ proofreading abilities and therefore requires a separate protein for this function. Alternatively, Cf114 (and the AgMNPV and CfDEFNPV homologues) may be involved in a DNA repair mechanism that acts autonomously from the DNA replication enzymes.

Repeated sequences

With the exception of CpGV, AdhoGV and NeleNPV, most baculoviruses contain *hr* regions composed of direct repeats with an imperfect palindromic 'core'. *hr*s have been implicated as origins of DNA replication in transient assays (Ahrens & Rohrmann, 1995; Kool *et al.*, 1995; Lee & Krell, 1994) and as enhancers of RNA polymerase II-mediated transcription (Theilmann & Stewart, 1992). We have identified five *hr*s in the CfMNPV genome, which is one more than reported earlier (Xie *et al.*, 1995) (Fig. 2). The number of repeats per *hr* ranged from seven in *hr1* and *hr3*, eight in *hr4*, to 10 in both *hr5* and *hr2* and accounted for 2·3 % of the genome. The five *hr*s were dispersed randomly in the genome with 3988 bp separating *hr1* and *hr2*, 42 710 bp separating *hr2* and *hr3*, 32 606 bp separating *hr3* and *hr4*, 21 327 bp separating *hr4* and *hr5* and 26 609 bp separating *hr5* and *hr1*. The multiple alignment of the CfMNPV *hr*s (Fig. 2a) showed a high degree of conservation within and between the *hr*s. The derived consensus repeat of the CfMNPV *hr*s, with an imperfect palindrome (shown in bold, Fig. 2b) demonstrated a high degree of similarity to the repeats found in the consensus OpMNPV *hr*s and to the consensus *hr* repeat from HycuNPV. Even though the *hr*s were dispersed randomly, the group I NPVs showed conserved genomic locations for certain *hr*s. *hr1* and *hr2* from CfMNPV were in the same relative genomic location as OpMNPV *hr2*, EppoNPV *hr2*, AcMNPV *hr2*, RoMNPV *hr2* and BmNPV *hr2L* and *2R* (near *fgf*). EppoNPV *hr5*, OpMNPV *hr5*, AcMNPV *hr1*, RoMNPV *hr1*, BmNPV *hr1* and CfMNPV *hr5* also shared the same relative genomic locations, between *odv-e56* and *ie-2* (Carstens *et al.*, 2002). Furthermore, all group I NPVs, with the exception of EppoNPV, had an *hr* near the *p95-cg30* loci. The relative location of CfMNPV *hr4* between *lef-7* and *chitinase* is identical to that of *hr4* of OpMNPV, while the *hr3* of EppoNPV is only about 1 kb away from *chitinase*. CfMNPV *hr4* is the only *hr* overlapping an ORF (Cf116), suggesting that this ORF may not be functional.

A high degree of variability existed around the *hr*s from CfMNPV, OpMNPV and EppoNPV (Fig. 3a–e). Among the group I NPVs, these regions contained gene deletions, unique genes in place of *hr*s or unique genes surrounding *hr*s. This was consistent with observations from MacoNPV A in which *bro* and *hr* sequences are in areas of variability between MacoNPV A and SeMNPV (Q. Li *et al.*, 2002). Hayakawa *et al.* (2000) also noted that *hr*s either flanked or were found within regions of major variations in gene content between AcMNPV and LdMNPV, and XcGV and AcMNPV.

Genomic organization, rearrangement and gene phylogeny

The gene order of CfMNPV was compared to those of other group I NPVs (AcMNPV, OpMNPV, CfDEFNPV), a group

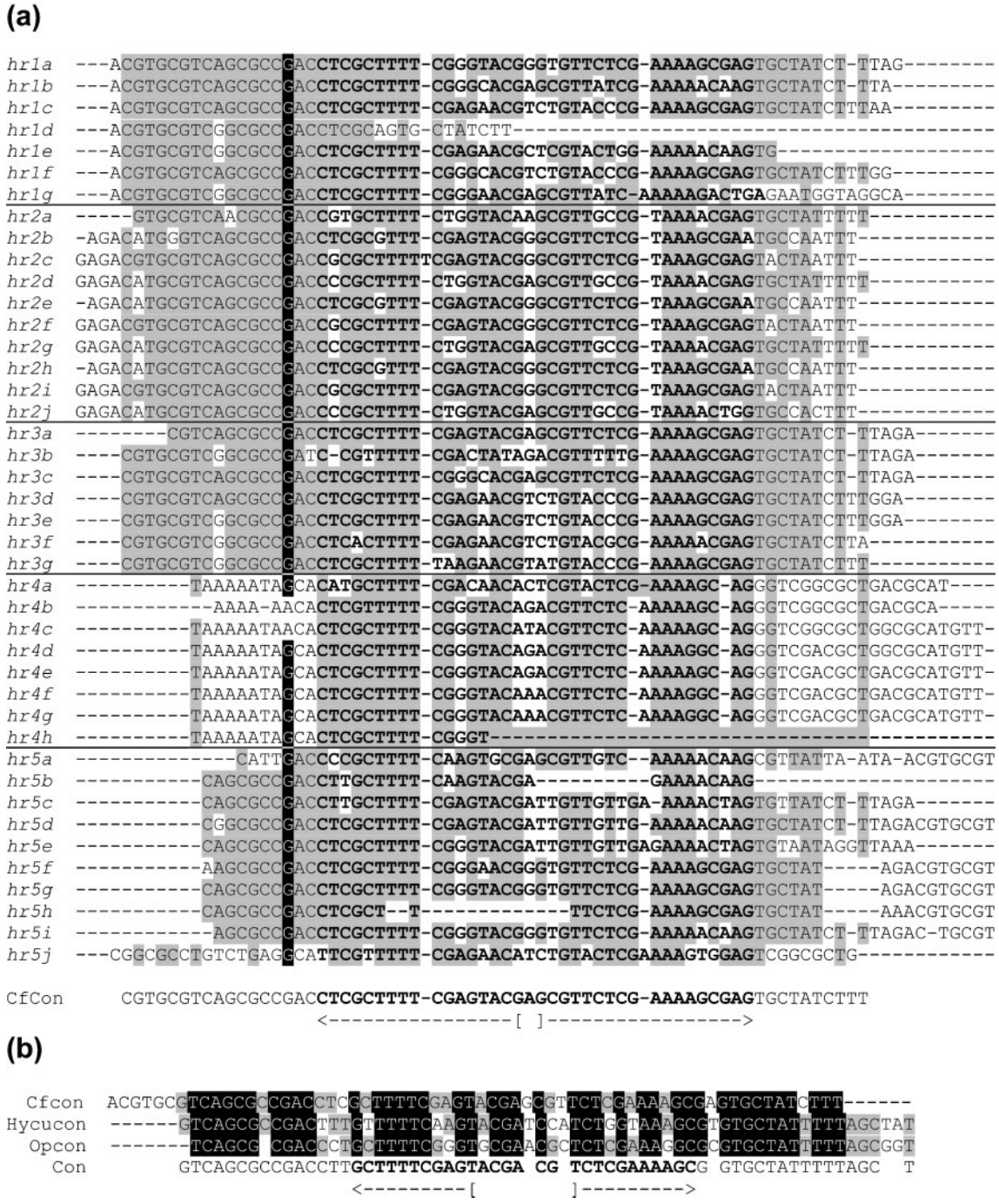

Fig. 2. (a) Alignment of CfMNPV *hr*s. White letters on black are highly conserved residues, black letters on grey are >50 % conserved, dashed lines (-) are gaps introduced to improve alignment. (b) Alignment of consensus *hr* repeats from CfMNPV, HycuNPV and OpMNPV. Bold letters indicate the palindrome nucleotides, bidirectional arrows indicate positions of the imperfect palindrome with [ ] indicating its central axis.

II NPV (HearNPV), a GV (CpGV) and a hymenopteran baculovirus (NeleNPV) by gene parity plots (Hu *et al.*, 1998) (Fig. 4). The organization of the CfMNPV genome was most collinear to that of OpMNPV, had two major areas of inversion compared with AcMNPV and one area of inversion compared with CfDEFNPV, a virus found in the same *C. fumiferana* host as CfMNPV. The inverted areas in AcMNPV were the same as those identified for OpMNPV

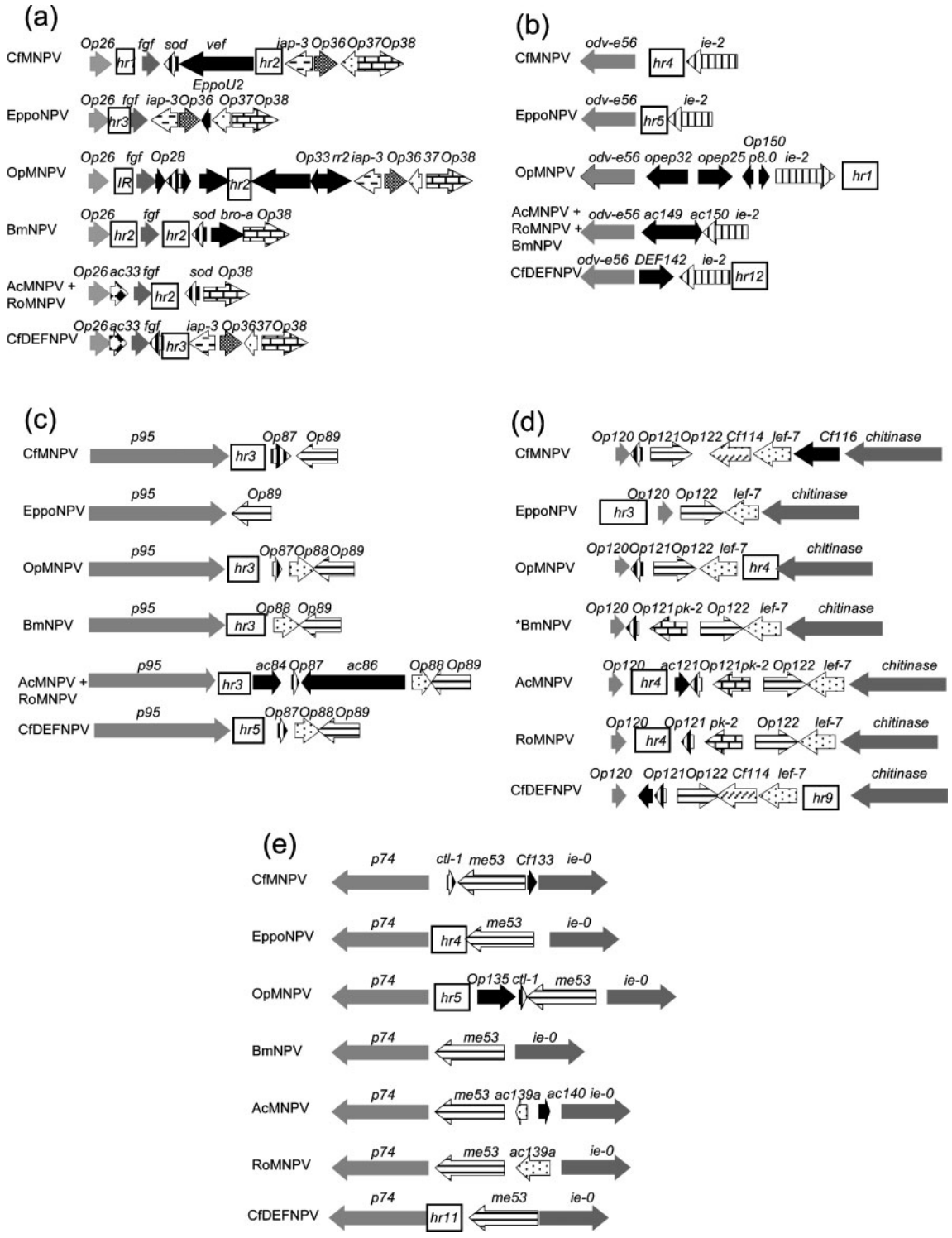

Fig. 3. Alignment of various *hr* regions from selected group I baculoviruses. Arrows with the same colour/pattern represent homologues, solid black arrows are ORFs found in only the one virus. The boxes represent *hr*s. Note: for this analysis AcMNPV, BmNPV and RoMNPV are equivalent where indicated. Asterisk (*) in (d) indicates an *hr* 2·5 kb upstream.

(Ahrens *et al.*, 1997; Lapointe *et al.*, 2000) involving *ac1–10* and *ac24–38*. When compared with a HearNPV G4, the genomes appeared less collinear, however conservation of gene order was noted in the 'central' region of the genomes. The gene parity plots for CfMNPV with CpGV and NeleNPV displayed a much more dispersed pattern.

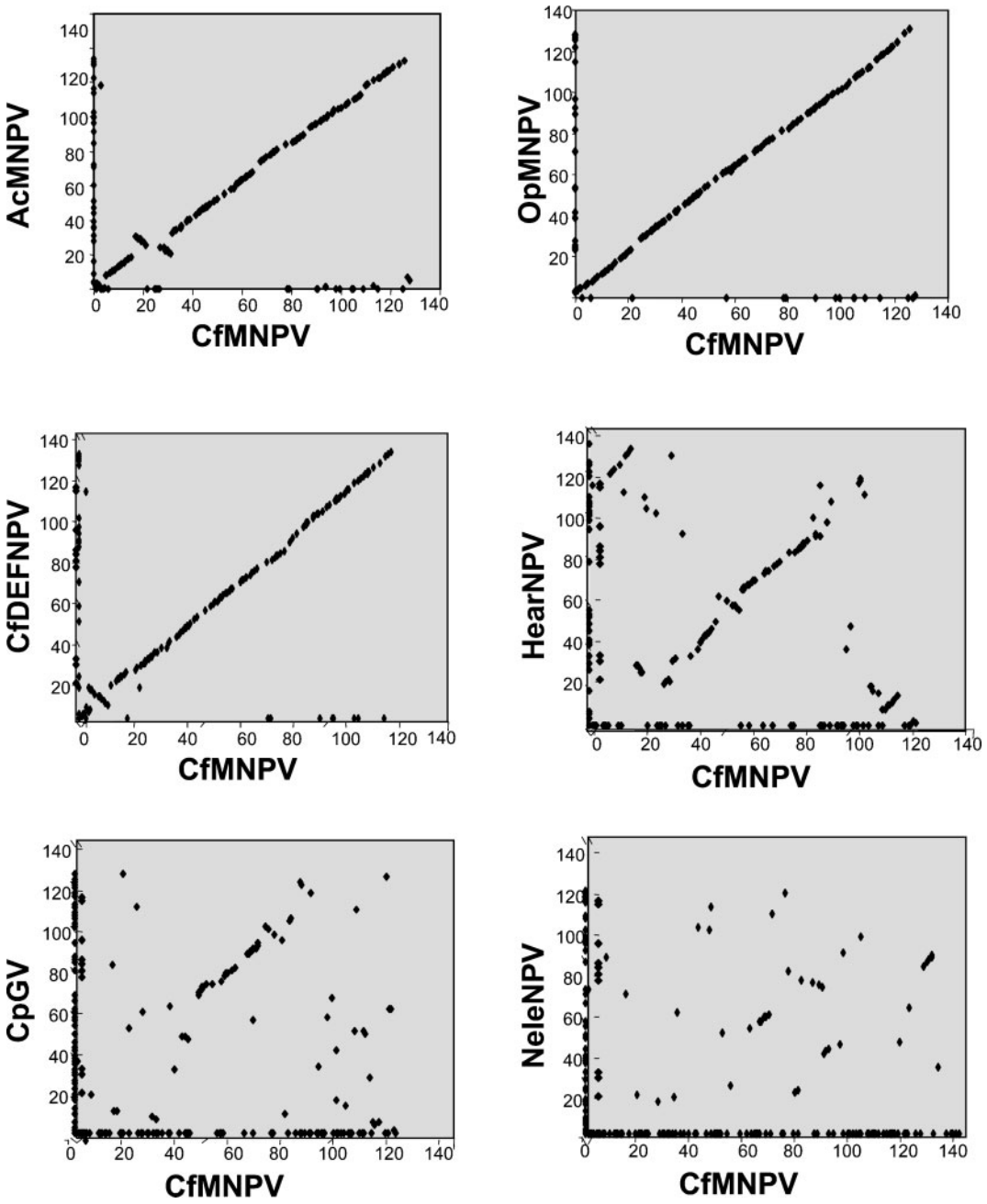

Fig. 4. Gene parity plot analysis of CfMNPV with AcMNPV, OpMNPV, CfDEFNPV, HearNPV G4, CpGV and NeleNPV, as indicated. The axes represent the relative position of each ORF along the genome in kb. The dots represent ORFs.

Based on the gene parity plot, phylogenetic analysis and overall amino acid identity among group I NPVs, CfMNPV appeared as a group I NPV most closely related to OpMNPV and to a lesser extent to EppoNPV, CfDEFNPV and AcMNPV.

Variations among group I NPV genomes

To date, a *vef* gene has been identified only in the GVs and type II NPVs (MacoNPV A, MacoNPV B and LdMNPV). VEFs are metalloproteinases that target the intestinal mucin of invertebrates leading to the degradation of the peritrophic membrane, thereby enhancing virus infectivity (Wang & Granados, 1997; Bischoff & Slavicek, 1997). CfMNPV is the first group I NPV to have a *vef* homologue. While multiple alignments revealed that CfMNPV VEF (Cf29) showed low similarity to other enhancing factors, a metalloproteinase zinc-binding signature domain (HEXXH) was identified. This signature pattern suggested that CfMNPV VEF may belong to the metalloproteinase superfamily (Jongeneel *et al.*, 1989).

That *vef* was not found in any other group I baculovirus suggested that CfMNPV acquired it through horizontal gene transfer. Furthermore, the genomic location of CfMNPV *vef* was in one of two regions of variability found among the type I baculoviruses identified by Hyink *et al.* (2002). Alignment of these genomic regions between *Op26* and *Op38*, or their corresponding homologues revealed a high degree of variation in ORF content between the AcMNPV, OpMNPV, EppoNPV, CfDEFNPV and CfMNPV genomes (Fig. 3a). Hyink *et al.* (2002) suggested that the common EppoNPV/OpMNPV ancestral virus at one point contained the *Op28–37* cluster and that EppoNPV lost *Op28–34* including *sod*. Considering that the EppoNPV/OpMNPV split occurred prior to the OpMNPV/CfMNPV split in the phylogenetic trees and that the CfMNPV *vef* falls within the *Op28–37* cluster in CfMNPV (Fig. 3a) *vef* might also have been part of this gene cluster in a shared ancestor. Alternatively, this region might have more readily acquired new genes through recombination with other viruses or hosts. Both OpMNPV and EppoNPV genomes have unique ORFs in this region, which may have been acquired in this way (Fig. 3a, *EppoU2* and *Op28*).

Hyink *et al.* (2002) identified a second region of variability around the *odv-e56* and *ie-2* gene loci suggesting that OpMNPV acquired the *Op147–150* cluster based on the orientation of *ie-2* which is reversed to that in AcMNPV and EppoNPV (Fig. 3b). The *Op147–150* cluster was not present in CfMNPV and the orientation of the CfMNPV *ie-2* gene was consistent with that in AcMNPV and EppoNPV thereby supporting the hypothesis presented by Hyink *et al.* (2002). Although CfMNPV lacked the *Op147–150* cluster, a CfMNPV unique gene (*Cf143*) was immediately downstream of *ie-2* (Fig. 1), and may have been incorporated due to a high rate of recombination in this area (Carstens *et al.*, 2002). Additionally, CfDEFNPV has a unique gene within this genomic region (Fig. 3b).

Alignment of several *hr* genomic regions from group I NPV further supports the theory of recombination within these areas (Fig. 3a–e). Within the genomic region encompassing *Op86–Op89*, CfMNPV lost *Op88*, while EppoNPV lacked *Op87*, *Op88* and the *hr* associated with this region. Additionally, AcMNPV retained *Op86* (*p95*) to *Op89*, the *hr* and has two genes, *ac84* and *ac86*, not found in OpMNPV, CfMNPV or EppoNPV (Fig. 3c). BmNPV, which is closely related to AcMNPV, had lost *Op87* and does not have *ac84* or *ac86*. CfDEFNPV shares the same genomic organization as OpMNPV for this region. The genomic region spanning *Op120* to *chitinase* also demonstrated variability (Fig. 3d). CfMNPV retained the *Op120–chitinase* cluster, however it lost the *hr* region and, instead, has two genes in its place, one of which is unique (*Cf114*, *Cf116*). EppoNPV lost a single gene, *Op121*, while AcMNPV retained the entire *Op120–chitinase* cluster, the *hr* and has two genes, *ac121* and *pk-2*, not found in CfMNPV, EppoNPV or OpMNPV. Both BmNPV and RoMNPV retained the *Op120–chitinase* cluster, as well as *pk-2*, however both lack *ac121*, and BmNPV lacks the *hr* in this region. Like CfMNPV, a *Cf114* homologue is found within this region in CfDEFNPV, however CfDEFNPV also acquired a unique gene between *Op120* and *Op121*. The final region shown (Fig. 3e) is the *p74–ie-0* genomic region in which EppoNPV, OpMNPV and CfDEFNPV have *hr*s, while CfMNPV, BmNPV, AcMNPV and RoMNPV do not. However, both CfMNPV and AcMNPV have genes (Fig. 3e, black arrows, *Cf133* and *ac140*) in this area not found in the other viruses used in this analysis. Moreover, despite both having retained *hr*s in this area, OpMNPV, EppoNPV and CfDEFNPV, differ as EppoNPV and CfDEFNPV have lost *Op135–ctl-1*. Given the high degree of variability within these selected regions, the *hr*s may play a key role in viral–viral and viral–host DNA recombination.

CfMNPV is also distinguished from other group I baculoviruses by the presence of two copies of *p26* (Cf7 and Cf128), a feature previously found only in the group II NPVs, MacoNPV A, MacoNPV B and SeMNPV. As in other group I NPVs, Cf128 (*p26b*) was immediately upstream of *p10*. In addition, Cf128 showed a relatively high amino acid identity (50·4–64·4 %) to other group I P26s while demonstrating lower identity to group II P26s. Conversely, Cf7 (*p26a*) was more homologous to group II *p26*s and less to group I *p26*s. This copy of *p26* was immediately upstream of *ptp-2* and downstream of a small ORF (68 aa) that demonstrated limited identity (69 % identity over 22 aa) to *ac2*, a *bro* gene. Considering the homology and genomic locations of the two copies of *p26*, it appears that Cf128 (*p26b*) may have been conserved from a common group I NPV ancestor, and that Cf7 (*p26a*) was acquired separately and not by gene duplication. The role of *p26* is unknown, however homologues for it have been identified in all NPVs sequenced to date except SpltMNPV (Pang *et al.*, 2001) and the two hymenopteran baculoviruses NeleNPV and NeseNPV (Lauzon *et al.*, 2004; Garcia-Maruniak *et al.*, 2004). Since one MacoNPV A *p26* had an early promoter and the second a late promoter in their respective upstream regions, Li & Rohrmann (2000) suggested that *p26* is required both early and late in infection and that having two copies fulfil this function. The CfMNPV *p26a* had a consensus early promoter within the first 150 bp, while *p26b* had both early and late promoters within the upstream 150 bp.

Interestingly a small remnant of a *bro* gene (*ac2*) was located adjacent to Cf7 (*p26a*), and *bro* genes have been implicated in recombination. In MacoNPV A, *bro* genes flanked regions of the genome containing unique genes or inversions and insertions when compared with other baculoviruses (Q. Li *et al.*, 2002). Kuzio *et al.* (1999) also noted that a region of the LdMNPV genome, which contained five contiguous *bro* sequences, was an area of frequent recombination. Given these observations, Cf7 may have been acquired through recombination via *bro*-like sequences.

As mentioned above, CfMNPV did not have an *iap-4* homologue as found in OpMNPV and EppoNPV. In these

genomes, *iap-4* occupies the same genomic location as *he65* in AcMNPV, RoMNPV and BmNPV, all of which lack *iap-4*. While CfMNPV lacked a full copy of *he65*, short contiguous sequences, in different reading frames, which displayed similarity to portions of *he65*, were found at the corresponding genomic location, accounting for much of the intergenic region between *p87* (Cf97) and Cf98 (Table 1). This suggested that a CfMNPV ancestor may have had a functional *he65* but had lost most of it, and what remained was rearranged and may be non-functional. Additionally, in OpMNPV and EppoNPV, which are quite closely related to CfMNPV, *iap-4* may have been acquired at the *he65* locus. Interestingly, in this genomic location in CfDEFNPV, there are two ORFs corresponding to the 3′ and 5′ ends of *he65* and a single copy of *iap-4*, adjacent to an *hr* element, further suggesting extensive recombination in this area.

In conclusion, based on concatenated protein phylogeny (Lauzon *et al.*, 2005) and gene parity plots, CfMNPV was most closely related to OpMNPV and to a lesser extent to EppoNPV. In phylogenetic trees, CfMNPV clusters with the group I NPVs and contains the entire group I-specific genes identified by Herniou *et al.* (2003). Despite its close relatedness to OpMNPV, the CfMNPV genome has several distinct features including the presence of a *vef* gene previously identified only in GVs and group II NPVs and contains two copies of the *p26* gene as found in several group II viruses. Herniou *et al.* (2003) proposed a possible model for the evolution of baculovirus gene content, whereby the ancestral lepidopteran virus acquired up to 45 genes beyond the 30 common baculovirus genes. Furthermore, the GVs and 14 of the lepidopteran NPVs had acquired 27 genes (Herniou *et al.*, 2003). The entire group I NPVs acquired an additional 12 genes (*ptp1*, *ac5*, *ac16*, *ac30*, *iap-1*, *gta*, *ac73*, *ac114*, *ac124*, *gp64/67*, *ac132* and *ie-2*). The AcMNPV, BmNPV and RoMNPV acquired five genes (*ac39*, *ac45*, *ac149*, *ac150* and *ac154*), while losing *iap-3* and the other group I NPV viruses acquired four genes (*Cf2*, *Cf31*, *Cf32* and *Cf104*) while losing *ac52*. Moreover, CfMNPV acquired *vef*, a second copy of *p26*, and the unique ORFs presented in Table 2, compared with all other group I viruses and either never acquired, or lost, *iap-4*, *Op4* and *Op106*, relative to OpMNPV. These differences identified between baculovirus genomes might provide insight into their evolution, host specificity and pathogenicity.

ACKNOWLEDGEMENTS

We would like to thank Angela Hollis from the Guelph Molecular Supercentre for her efforts in sequencing. J. G. D. was a recipient of an NSERC PGS-A scholarship. This work was funded by an NSERC strategic grant, an NSERC Biocontrol Network grant and Genome Canada/Ontario Genomics Institute grant to P. J. K. and B. M. A. and by grants from CIHR and Ontario Genomics Institute to E. B. C.

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. & Rohrmann, G. F. (1995). Replication of *Orgyia pseudotsugata* baculovirus DNA: *lef-2* and *ie-1* are essential and *ie-2*, *p34*, and Op-*iap* are stimulatory genes. *Virology* **212**, 650–662.

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.

Arif, B. M., Kuzio, J., Faulkner, P. & Doerfler, W. (1984). The genome of *Choristoneura fumiferana* nuclear polyhedrosis virus: molecular cloning and mapping of the *Eco*RI, *Bam*HI, *Sma*I, *Xba*I and *Bgl*II restriction sites. *Virus Res* **1**, 605–614.

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.

Birnbaum, M. J., Clem, R. J. & Miller, L. K. (1994). An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. *J Virol* **68**, 2521–2528.

Bischoff, D. S. & Slavicek, J. M. (1997). Molecular analysis of an *enhancin* gene in the *Lymantria dispar* nuclear polyhedrosis virus. *J Virol* **71**, 8133–8140.

Blissard, G. W., Black, B., Crook, N., Keddie, B. A., Possee, R., Rohrmann, G., Theilmann, D. & Volkman, L. (2000). Family *Baculoviridae*. In *Virus Taxonomy. The Seventh Report of the International Committee on Taxonomy of Viruses*, pp. 195–202. Edited by M. H. 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. San Diego: Academic Press.

Carstens, E. B., Liu, J. J. & Dominy, C. (2002). Identification and molecular characterization of the baculovirus *Cf*MNPV early genes: *ie-1*, *ie-2* and *pe38*. *Virus Res* **83**, 13–30.

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.

Chen, T., Sahri, D. & Carstens, E. B. (2004). Characterization of the interaction between P143 and LEF-3 from two different baculovirus species: *Choristoneura fumiferana* nucleopolyhedrovirus LEF-3 can complement *Autographa californica* nucleopolyhedrovirus LEF-3 in supporting DNA replication. *J Virol* **78**, 329–339.

Crook, N. E., Clem, R. J. & Miller, L. K. (1993). An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *J Virol* **67**, 2168–2174.

Garcia-Maruniak, A., Maruniak, J. E., Zanotto, P. M., Doumbouya, A. E., Liu, J. C., Merritt, T. M. & Lanoie, J. S. (2004). Sequence analysis of the genome of the *Neodiprion sertifer* nucleopolyhedrovirus. *J Virol* **78**, 7036–7051.

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

Guarino, L. A., Xu, B., Jin, J. & Dong, W. (1998). A virus-encoded RNA polymerase purified from baculovirus-infected cells. *J Virol* **72**, 7985–7991.

Hang, X., Dong, W. & Guarino, L. A. (1995). The *lef-3* gene of *Autographa californica* nuclear polyhedrosis virus encodes a single-stranded DNA-binding protein. *J Virol* **69**, 3924–3928.

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. I., Goto, C. & Maeda, S. (1999). Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* **262**, 277–297.

Hayakawa, T., Rohrmann, G. F. & Hashimoto, Y. (2000). Patterns of genome organization and content in lepidopteran baculoviruses. *Virology* **278**, 1–12.

Herniou, E. A., Olszewski, J. A., Cory, J. S. & O'Reilly, D. R. (2003). The genome sequence and evolution of baculoviruses. *Annu Rev Entomol* **48**, 211–234.

Hill, J. E., Kuzio, J., Wilson, J. A., MacKinnon, E. A. & Faulkner, P. (1993). Nucleotide sequence of the *p74* gene of a baculovirus pathogenic to the spruce budworm, *Choristoneura fumiferana* multicapsid nuclear polyhedrosis virus. *Biochim Biophys Acta* **1172**, 187–189.

Hill, J. E., Kuzio, J. & Faulkner, P. (1995). Identification and characterization of the *v-cath* gene of the baculovirus, *Cf*MNPV. *Biochim Biophys Acta* **1264**, 275–278.

Hu, Z. H., Arif, B. M., Jin, F., Martens, J. W., Chen, X. W., Sun, J. S., Zuidema, D., Goldbach, R. W. & Vlak, J. M. (1998). Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. *J Gen Virol* **79**, 2841–2851.

Huang, J. & Levin, D. B. (2001). Expression, purification and characterization of the *Spodoptera littoralis* nucleopolyhedrovirus (SpliNPV) DNA polymerase and interaction with the SpliNPV non-*hr* origin of DNA replication. *J Gen Virol* **82**, 1767–1776.

Hyink, O., Dellow, R. A., Olsen, M. J., Caradoc-Davies, K. M., 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., van Strien, E. A., Heldens, J. G., 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.

Jongeneel, C. V., Bouvier, J. & Bairoch, A. (1989). A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS Lett* **242**, 211–214.

Kool, M., Ahren, C. H., Goldbach, R. W., Rohrmann, G. F. & Vlak, J. M. (1994). Identification of genes involved in DNA replication of the *Autographa californica* baculovirus. *Proc Natl Acad Sci U S A* **91**, 11212–11216.

Kool, M., Ahrens, C. H., Vlak, J. M. & Rohrmann, G. F. (1995). Replication of baculovirus DNA. *J Gen Virol* **76**, 2103–2118.

Kost, T. A. & Condreay, J. P. (1999). Recombinant baculoviruses as expression vectors for insect and mammalian cells. *Curr Opin Biotechnol* **10**, 428–433.

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.

Lapointe, R., Back, D. W., Ding, Q. & Carstens, E. B. (2000). Identification and molecular characterization of the *Choristoneura fumiferana* multicapsid nucleopolyhedrovirus genomic region encoding the regulatory genes *pkip*, *p47*, *lef-12*, and *gta*. *Virology* **271**, 109–121.

Lauzon, H. A. M., Lucarotti, C. J., Krell, P. J., Feng, Q., Retnakaran, A. & Arif, B. M. (2004). Sequence and organization of the *Neodiprion lecontei* nucleopolyhedrovirus genome. *J Virol* **78**, 7023–7035.

Lauzon, H. A. M., Jamieson, P. B., Krell, P. J. K. & Arif, B. M. (2005). Gene organization and sequencing of the *Choristoneura fumiferana* defective nucleopolyhedrovirus genome. *J Gen Virol* **86**, 945–961.

Lee, H. & Krell, P. J. (1994). Reiterated DNA fragments in defective genomes of *Autographa californica* nuclear polyhedrosis virus are competent for AcMNPV-dependent DNA replication. *Virology* **202**, 418–429.

Li, L. & Rohrmann, G. F. (2000). Characterization of a baculovirus alkaline nuclease. *J Virol* **74**, 6401–6407.

Li, X., Pang, A., Lauzon, H. A., Sohi, S. S. & Arif, B. M. (1997). The gene encoding the capsid protein P82 of the *Choristoneura fumiferana* multicapsid nucleopolyhedrovirus: sequencing, transcription and characterization by immunoblot analysis. *J Gen Virol* **78**, 2665–2673.

Li, X., Lauzon, H. A., Sohi, S. S., Palli, S. R., Retnakaran, A. & Arif, B. M. (1999). Molecular analysis of the p48 gene of *Choristoneura fumiferana* multicapsid nucleopolyhedroviruses CfMNPV and CfDEFNPV. *J Gen Virol* **80**, 1833–1840.

Li, L. & Rohrmann, G.F. (2000). Characterization of a baculovirus alkaline nuclease. *J Virol* **74**, 6401–6407.

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

Liu, J. J. & Carstens, E. B. (1993). Infection of *Spodoptera frugiperda* and *Choristoneura fumiferana* cell lines with the baculovirus *Choristoneura fumiferana* nuclear polyhderosis virus. *Can J Microbiol* **39**, 932–939.

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.

Liu, J. J. & Carstens, E. B. (1996). Identification, molecular cloning, and transcription analysis of the *Choristoneura fumiferana* nuclear polyhedrosis virus spindle-like protein gene. *Virology* **223**, 396–400.

Lu, A. & Miller, L. K. (1995). The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *J Virol* **69**, 975–982.

Luque, T., 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.

Maguire, T., Harrison, P., Hyink, O., Kalmakoff, J. & Ward, V. K. (2000). The inhibitors of apoptosis of *Epiphyas postvittana* nucleopolyhedrovirus. *J Gen Virol* **81**, 2803–2811.

Mikhailov, V. S., Mikhailova, A. L., Iwanaga, M., Gomi, S. & Maeda, S. (1998). *Bombyx mori* nucleopolyhedrovirus encodes a DNA-binding protein capable of destabilizing duplex DNA. *J Virol* **72**, 3107–3116.

Mikhailov, V. S., Okano, K. & Rohrmann, G. F. (2003). Baculovirus alkaline nuclease possesses a 5′→3′ exonuclease activity and associates with the DNA-binding protein LEF-3. *J Virol* **77**, 2436–2444.

Miller, L. K. (1997). Introduction to the Baculoviruses. In *The Baculoviruses*, pp. 1–6. Edited by L. K. Miller. New York: Plenum.

Monsma, S. A., Oomens, A. G. & Blissard, G. W. (1996). The GP64 envelope fusion protein is an essential baculovirus protein required for cell-to-cell transmission of infection. *J Virol* **70**, 4607–4616.

Nakai, M., Goto, C., Kang, W., 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.

Pang, Y., Yu, J., Wang, L. & 7 other authors (2001). Sequence analysis of the *Spodoptera litura* multicapsid nucleopolyhedrovirus genome. *Virology* **287**, 391–404.

Pearson, M. N., Groten, C. & Rohrmann, G. F. (2000). Identification of the *Lymantria dispar* nucleopolyhedrovirus envelope fusion protein provides evidence for a phylogenetic division of the *Baculoviridae*. *J Virol* **74**, 6126–6131.

Slack, J. M., Ribeiro, B. M. & de Souza, M. L. (2004). The *gp64* locus of *Anticarsia gemmatalis* multicapsid nucleopolyhedrovirus contains a 3′ repair exonuclease homologue and lacks *v-cath* and *ChiA* genes. *J Gen Virol* **85**, 211–219.

Theilmann, D. A. & Stewart, S. (1992). Tandemly repeated sequence at the 3′ end of the IE-2 gene of the baculovirus *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus is an enhancer element. *Virology* **187**, 97–106.

Wang, P. & Granados, R. R. (1997). An intestinal mucin is the target substrate for a baculovirus enhancin. *Proc Natl Acad Sci U S A* **94**, 6977–6982.

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.

Wormleaton, S., Kuzio, J. & Winstanley, D. (2003). The complete sequence of the *Adoxophyes orana* granulovirus genome. *Virology* **311**, 350–365.

Wu, Y. & Carstens, E. B. (1998). A baculovirus single-stranded DNA binding protein, LEF-3, mediates the nuclear localization of the putative helicase P143. *Virology* **247**, 32–40.

Xie, W. D., Arif, B., Dobos, P. & Krell, P. J. (1995). Identification and analysis of a putative origin of DNA replication in the *Choristoneura fumiferana* multinucleocapsid nuclear polyhedrosis virus genome. *Virology* **209**, 409–419.

Yang, S. & Miller, L. K. (1998). Control of baculovirus polyhedrin gene expression by very late factor 1. *Virology* **248**, 131–138.

Yang, S. & Miller, L. K. (1999). Activation of baculovirus very late promoters by interaction with very late factor 1. *J Virol* **73**, 3404–3409.

Yang, D.-H., de Jong, J. G., Makhmoudova, A., Arif, B. M. & Krell, P. J. (2004). *Choristoneura fumiferana* nucleopolyhedrovirus encodes a functional 3′-5′ exonuclease. *J Gen Virol* **85**, 3569–3573.