Journal of General Virology (2006), **87**, 1477–1489 DOI 10.1099/vir.0.81727-0

Genomic comparison of *Neodiprion sertifer* and *Neodiprion lecontei* nucleopolyhedroviruses and identification of potential hymenopteran baculovirus-specific open reading frames

Hilary A. M. Lauzon,[1]† Alejandra Garcia-Maruniak,[2]† Paolo M. de A. Zanotto,[3] José C. Clemente,[4] Elisabeth A. Herniou,[5] Christopher J. Lucarotti,[6] Basil M. Arif[1] and James E. Maruniak[2]

Correspondence
James E. Maruniak
jem@ifas.ufl.edu

[1]Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste Marie, ON P6A 2E5, Canada

[2]Entomology and Nematology Department, University of Florida, Bldg 970, Natural Area Drive, Gainesville, FL 32611, USA

[3]LEMB, Instituto de Ciências Biomédicas, USP, Av. Lineu Prestes 1374, CEP 05508-900, São Paulo, SP, Brazil

[4]Department of Biochemistry and Molecular Biology, University of Florida College of Medicine, PO Box 100245, Gainesville, FL 32610-0245, USA

[5]Division of Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK

[6]Canadian Forest Service, Atlantic Forestry Centre, PO Box 4000, Regent Street, Fredericton, NB E3B 5P7, Canada

Genomic comparison of *Neodiprion sertifer nucleopolyhedrovirus* (NeseNPV) and *Neodiprion lecontei nucleopolyhedrovirus* (NeleNPV) showed that the hymenopteran baculoviruses had features in common and were distinct from other, fully sequenced lepidopteran and dipteran baculoviruses. Their genomes were small in size (86 462 and 81 755 bp, respectively), had low G+C contents (33·8 and 33·3 mol%, respectively) and contained fewer open reading frames (ORFs) (90 and 89, respectively) than other baculoviruses. They shared 69 ORFs (48·6 % mean amino acid identity overall), 43 of which were previously identified baculovirus homologues. The remaining shared ORFs could be common to other baculoviruses, but low amino acid identities precluded identifying them as such. Some may also be unique to hymenopteran baculoviruses. These included a trypsin-like protease, a zinc-finger protein, regulator of chromosome condensation proteins, a densovirus capsid-like protein and a phosphotransferase. Structural analysis, the presence of conserved domains and phylogenetic studies suggested that some of these ORFs may be functional and could have been transferred horizontally from an insect host. ORFs found only in NeseNPV and NeleNPV may play a role in host specificity and/or tissue tropism, as hymenopteran baculoviruses are restricted to the midgut. The genomes were basically collinear, but contained non-syntenic regions (NSRs) with large numbers of repeats between their *polyhedrin* and *dbp* genes. They differed from each other in the number of ORFs and the G+C content of their NSRs and the presence of homologous regions in the NeseNPV genome. NeleNPV also had a short inversion relative to NeseNPV. NeseNPV contained 21 ORFs not found in NeleNPV and NeleNPV had 20 ORFs not found in NeseNPV.

Received 30 November 2005
Accepted 31 January 2006

INTRODUCTION

The majority of baculoviruses infect insects from the order Lepidoptera, but they have also been isolated from members of Diptera, Hymenoptera and a limited number of other hosts (Volkman *et al.*, 1995). Until recently, most fully sequenced baculovirus genomes were from lepidopteran hosts and were divided into either group I or II nucleopolyhedroviruses (NPVs) or granuloviruses (GVs) (Herniou *et al.*, 2001). Sequencing of the *Culex nigripalpus* NPV (CuniNPV) genome showed that the dipteran baculoviruses are at a large evolutionary distance from the lepidopteran NPVs and

†These authors contributed equally to this work.

GVs (Afonso *et al.*, 2001; Herniou *et al.*, 2003). Partial sequencing of the *lef*-8 and ac22 homologues from hymenopteran baculoviruses *Neodiprion sertifer* NPV (NeseNPV), *Neodiprion lecontei* NPV (NeleNPV) and *Gilpinia hercyniae* NPV similarly showed a division of the baculoviruses according to the order of their hosts (Herniou *et al.*, 2004). The recent sequencing of the NeseNPV and NeleNPV genomes has shown clearly that the hymenopteran viruses are evolutionarily distant from the lepidopteran and dipteran baculoviruses and may belong to a novel genus of the family *Baculoviridae* (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004).

NeseNPV infects the European pine sawfly, *N. sertifer*, a serious defoliator of pine in many parts of Europe and Asia. This Old World insect was unknown in North America until its introduction to the US in 1925 and to Canada in 1939 (Brown, 1982), where it became an important pest of coniferous forests. NeseNPV was introduced to North America from Sweden in 1950 as a biological-control agent against *N. sertifer* (Bird, 1953; Brown, 1982), later becoming available as a registered product called Neochek-S (Huber, 1986). NeleNPV was first identified in Ontario in 1950 in the red-headed pine sawfly, *N. lecontei* (Bird, 1961). *N. lecontei* attacks young, natural pine stands, plantations and greenhouse cultures, causing defoliation, death of young trees, reduced growth and tree deformity (Cunningham *et al.*, 1984). NeleNPV is also available as a registered product called Lecontvirus and has been used successfully for many years as a biological-control agent against *N. lecontei* (De Groot & Cunningham, 1983; http://www.glfc.cfs.nrcan.gc.ca/Lecontvirus.pdf).

Unlike the majority of lepidopteran baculoviruses, hymenopteran baculoviruses replicate only in the epithelial cells of the larval midgut. The gregarious nature of many sawflies, combined with the excretion of infective virus from the midgut prior to insect death, leads to the rapid spread of the viruses with insects, dying 4–7 days after infection (Federici, 1997). Dipteran baculoviruses are also restricted to midgut replication, but their feeding ecologies differ from those of sawflies. Hymenopteran and lepidopteran larvae feed on terrestrial plants, whereas mosquito larvae are aquatic (Afonso *et al.*, 2001; Herniou *et al.*, 2004).

NeseNPV and NeleNPV have smaller genomes than other sequenced baculoviruses, contain fewer ORFs and share limited similarity with lepidopteran NPVs, GVs and CuniNPV. NeseNPV and NeleNPV were related more closely to each other than to other baculoviruses, but even though they infect hosts from the same genus, they did not share as high a degree of sequence identity with each other as with some lepidopteran baculoviruses infecting hosts from different families (Lauzon *et al.*, 2005). This paper compares the genomes of NeseNPV and NeleNPV and discusses their similarities and differences.

METHODS

Comparison of *Neodiprion* baculoviruses. Complete protein alignments from translated homologues identified in NeseNPV (Garcia-Maruniak *et al.*, 2004) and NeleNPV (Lauzon *et al.*, 2004) were done by using CLUSTAL W from Lasergene's DNAStar MEGALIGN program (version 5.06) with default conditions. All remaining open reading frames (ORFs) were compared with those in GenBank by using the National Center for Biotechnology Information (NCBI) BLAST searches (Altschul *et al.*, 1990, 1997), including standard protein–protein BLAST (BLASTP) searches. The Viral Orthologous Clusters (VOCs) program from the Viral Bioinformatics Resource Center (www.virology.ca) was also used with default settings to identify further homologues. New potential NeseNPV and NeleNPV matches were considered homologous if amino acid identity of complete ORFs was >20 %. If amino acid identity of proteins was <20 %, matches were noted, but were not included when calculating overall mean amino acid identities or in homologue totals. A syntenic map was prepared, in order to compare the collinearity of the genomes. The genomes were aligned with the BLAST2 sequence tool by using the TBLASTX program. They were translated in all six reading frames and the encoded peptides of one genome were compared with those of the other. The coordinates of all high-similarity pairs (HSPs) in the output were then used with the Artemis Comparative Tool (ACT) (http://www.sanger.ac.uk/Software/Artemis/) to generate a syntenic map showing the collinearity of conserved regions between genomes (not specifically ORFs). The HSP scores accepted for this analysis were from 27 to 405, with identities ranging from 27 to 67 %. ORFs were analysed with the Simple Modular Architectural Research Tool (SMART) (Schultz *et al.*, 1998, 2000). Conserved protein domains were aligned by using CLUSTAL W and shading was done with GeneDoc. Phylogenetic reconstructions were done by using maximum likelihood with PHYML (global tree search) (http://www.lirmm.fr/mab/sommaire_english.php3) (Guindon & Gascuel, 2003), TREE-PUZZLE 5.2 (quartet method) (http://www.tree-puzzle.de/) and global parsimony by heuristic search using PAUP 4.0b10 (Swofford, 2003). Support for trees was obtained after 1000 non-parametric bootstrap replications with PHYML, by the number of quartets supporting a given node with TREE-PUZZLE and after 1000 bootstrap replications with PAUP. Only nodes with support above 50 % were annotated.

NeseNPV and NeleNPV trypsin-like serine protease three-dimensional (3D) modelling. The 3D structures of the trypsin-like serine proteases from NeseNPV ORF 7 (nese7) and NeleNPV ORF 6 (nele6) were modelled by structural-homology methods utilizing the COMPOSER module in SYBYL 7.0 (Tripos Inc.). The pro-segment regions were removed and the sequences from aa 30 for nese7 and aa 31 for nele6 were used for modelling. Disulfide bonds were built manually based on sequence homology and known trypsin structures (nese7: Cys55–Cys71, Cys182–Cys198 and Cys209–Cys233; nele6: Cys56–Cys72, Cys183–Cys199 and Cys210–Cys234). The proteases were minimized by using the Powell method and Tripos force field to an energy gradient of 0·05 kcal mol^{-1} $Å^{-1}$. Gasteiger–Hückel charges were used in the minimization (Purcell & Singer, 1967). Structural overlaps and figures were generated by using PyMOL (v0.98; Molecular Graphics System) (DeLano, 2002). Default settings were used with all programs.

RESULTS AND DISCUSSION

Comparison of *Neodiprion* baculoviruses

NeseNPV and NeleNPV are the smallest baculovirus genomes sequenced to date, with sizes of 86 462 bp (90 ORFs) and 81 755 bp (89 ORFs), respectively. Both genomes have low G+C contents: 33·8 mol% for NeseNPV and 33·3 mol% for NeleNPV (Table 1). Based on analysis of the individual genomes, each virus was reported to have 43 baculovirus homologues (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*,

Table 1. Characteristics of the NeseNPV and NeleNPV genomes

Characteristic	NeseNPV	NeleNPV
Size (bp)	86 462	81 755
G+C content (mol%)	33·8	33·3
Total no. ORFs	90	89
No. *hr*s	6	0
No. *dr*s	4	9
NeseNPV/NeleNPV shared ORFs	69	69
Shared ORFs found only in NeseNPV and NeleNPV	26	26
Shared ORFs found in other baculoviruses	43	43
Non-shared ORFs found in other baculoviruses	1	1
Non-shared ORFs not found in other baculoviruses	20	19

2004). The NeseNPV genome contained a methyltransferase (*mtase1*) homologue (nese5/ac69) not reported in NeleNPV and NeleNPV had an ac76 homologue (nele41) not reported in NeseNPV. Comparative analysis showed 58·4 % amino acid identity between nele41 and nese44, inferring that nese44 is an ac76 homologue; however, no direct match was obtained between nese44 and ac76 homologues. No homologue of nese5 was found in NeleNPV, thus the latter still lacks an identified *mtase1*. The VOCs program identified nele79 as a *Xestia-c-nigrum* granulovirus ORF 164 homologue (20·4 % amino acid identity), but no homologue was found in NeseNPV. By adding the inferred NeseNPV ac76 homologue and the xecn164 homologue to NeleNPV, each genome now has 44 baculovirus homologues, with 43 of them shared with each other.

CLUSTAL W analysis of total protein sequences indicated that 69 ORFs were common to NeseNPV and NeleNPV, 26 of which did not appear to match previously identified baculovirus homologues. Some ORFs found only in NeleNPV and NeseNPV may be baculoviral homologues with identities too low to be identified clearly, or they may be specific to hymenopteran baculoviruses. Analysis indicated that a few other NeleNPV ORFs were potentially shared with NeseNPV. nele4 and nele8, for example, had potential matches to nese9, nese11, nese12, nese18 and nese19 and to each other, but only nele4/nese9 had an amino acid identity >20 %. The multiple matches of nele4/nele8 with NeseNPV ORFs suggest a duplication of ORFs. nele13 also had a significant BLASTP match with nese20, but the amino acid identity was only 17·7 %. Potential matches are noted under comments in Table 2, but were not included in homologue totals or amino acid means. A previously unreported similarity between nele25 and nele26 (25·6 % amino acid identity) was found, but neither ORF had homologues in NeseNPV. Counting only the ORFs with >20 % amino acid identity, NeseNPV had 21 ORFs that were not found in NeleNPV and NeleNPV had 20 ORFs not found in NeseNPV (Tables 1 and 2).

The mean amino acid identity for the 69 ORFs common to NeseNPV and NeleNPV was 48·6 %. This was higher than the mean amino acid identity reported for shared ORFs between NeleNPV and the dipteran virus CuniNPV (19·7 % amino acid identity) and between NeleNPV and various lepidopteran baculoviruses, which ranged from 23·2 % with *Autographa californica* MNPV (AcMNPV) to 24·0 % with *Spodoptera exigua* MNPV (Lauzon *et al.*, 2004), but was lower than the mean amino acid identity shared between many lepidopteran baculovirus genomes (Lauzon *et al.*, 2005). The mean amino acid identity for the 43 ORFS shared by NeseNPV and NeleNPV and present in other baculoviruses was 57·9 %, but for the 26 ORFs found only in NeleNPV and NeseNPV, it was 41·5 %, suggesting that there might be more selection pressure on the ORFs shared with other baculoviruses. Overall, the ORFs with the highest amino acid identity in NeseNPV and NeleNPV were ODV-E18 (84·5 %), POLYHEDRIN (82·1 %) and PIF-2 (74·3 %) (Table 2).

Non-syntenic regions (NSRs)

The most obvious differences between the two genomes were found in the regions between the *polyhedrin* and DNA-binding protein (*dbp*) genes (nese1–nese22 and nele1–nele14) (Table 2). Due to the lack of conserved synteny in these areas, they have been called NSRs. The NSRs are clearly seen in the syntenic map that compares the conserved regions between the two linearized genomes (Fig. 1). In the NSRs, NeseNPV had 15 ORFs that were not found in NeleNPV and NeleNPV had eight ORFs not found in NeseNPV. The discrepancy in the number of ORFs may be partially due to the duplication of ORFs. nese18 and nese19 were previously considered duplicate genes because they shared 71·2 % amino acid identity, were in the opposite orientation, lacked identifiable upstream promoters and each ORF was flanked by direct repeats (*dr*s) or homologous regions (*hr*s) (Garcia-Maruniak *et al.*, 2004). nele4 and nele8 had tentative matches to nese9, nese11, nese12, nese18 and nese19. The only previously known baculovirus homologues in the NSRs were the *mtase1* homologue in NeseNPV (nese5/ac69) and the *iap* homologues (nele11/nese17). It is noteworthy, however, that the closest BLASTP matches for these ORFs were to insect proteins. The highest match for nese5 was to a honeybee (*Apis mellifera*) protein (GenBank accession no. XP_394722). The highest matches for nele11 were an *A. mellifera* SON DNA-binding protein (GenBank accession no. XP_396370) and an inhibitor of apoptosis (IAP) from *Spodoptera frugiperda* (GenBank accession no. AAF35285), and for nese17, a *Bombyx mori* IAP (GenBank accession no. AAK57560). Other ORFs in the NSRs, such as nele6/nese7 (trypsin-like serine proteases), also showed closest matches to insect proteins. It is conceivable that the NSRs arose by horizontal transfer of a gene cluster from an insect host(s) and only genes useful to the virus had enough selection pressure to be maintained or to prevent extensive mutations. This might account for the different ORF

Table 2. Comparison of NeseNPV and NeleNPV ORFs

hr, Homologous region; *dr*, direct repeat; DSRM, double-stranded RNA-binding motif; SP, signal-peptide motif; TMD, transmembrane motif. Only baculovirus homologues with amino acid identity of 20 % or more are shown under the 'Identity (%)' column. No. ORFs shared by NeseNPV and NeleNPV, 69. Bold type indicates shared ORFs with homology to other baculoviruses ($n=43$); italic type indicates shared ORFs with no homology to other baculoviruses ($n=26$).

Name	NeleNPV		NeseNPV			Baculovirus ORF*	Comments†
	ORF	Length (aa)	ORF	Length (aa)	Identity (%)		
polh	**1**	**247**	**1**	**246**	**82·1**	**ac8**	
	2	108					
			2	195			Internal repeats
	3	80					
			4	56			
mtase1			5	437		ac69	
			6	113			
			8	202			
	4	*332*	*9*	*416*	*22·5*		nese11, nese12, nese18, nese19
	5	52					
	6	*259*	*7*	*258*	*72·6*		Trypsin-like
	7	131					Internal repeats
NeleNPV *dr*1							2 × 43 bp
	8	526					Internal repeats, nele4, nese9, nese11, nese12, nese18, nese19
			10	56			
NeseNPV *hr*1							5 × 65 bp; 4 × 45 bp
			11	461			
NeseNPV *hr*2							12 × 65 bp
			12	130			
			13	146			
			14	58			
			15	54			
NeseNPV *dr*1 inside nese16	*9*	*794*	*16*	*637*	*29·0*		2 × 120 bp (in nese16) Ac66 desmoplakin
	10	54					
iap-3	**11**	**260**	**17**	**181**	**24·2**	**op35**	
NeseNPV *hr*3							7 × 65 bp; 7 × 45 bp
			18	301			Duplicate nese19
NeseNPV *dr*2							3 × 29 bp
NeseNPV *hr*4							7 × 65 bp; 6 × 45 bp
			19	307			Duplicate nese18
NeseNPV *hr*5							9 × 45 bp
			20	270			Nele13 17·7 % amino acid identity
			21	89			
	12	80					
	13	247					Nese20 17·7 % amino acid identity
dbp	**14**	**236**	**22**	**251**	**44·7**	**ac25**	
lef-11	**15**	**101**	**23**	**100**	**38·6**	**ac37**	
p33	**16**	**256**	**24**	**252**	**50·2**	**ac92**	
p18	**17**	**170**	**25**	**170**	**47·4**	**ac93**	
	18	*216*	*26*	*218*	*41·9*		
			27	57			
	19	63					
dna pol	**20**	**923**	**28**	**913**	**61·4**	**ac65**	
	21	*728*	*29*	*792*	*28·9*		ac66 desmoplakin
	22	*380*	*3*	*388*	*23·5*		

Table 2. cont.

Name	NeleNPV		NeseNPV			Baculovirus ORF*	Comments†
	ORF	Length (aa)	ORF	Length (aa)	Identity (%)		
odv-e56	**23**	**336**	**38**	**339**	**71·2**	**ac148**	Inversion begins
	24	*192*	*39*	*192*	*50·3*		
	25	261					
NeleNPV *dr2*							2 × 53 bp
	26	226					
	27	*190*	*37*	*133*	*21·6*		
p6.9	**28**	**102**	**36**	**86**	**71·3**	**ac100**	
c42/p40	**29**	**366**	**35**	**368**	**63·5**	**ac101**	
	30	*116*	*34*	*117*	*47·9*		
p48	**31**	**390**	**33**	**402**	**55·2**	**ac103**	
	32	**238**	**32**	**235**	**61·9**	**ac106/107**	
alk-exo	**33**	**402**	**31**	**399**	**48·0**	**ac133**	
			30	78			Inversion ends
	34	157					
	35	97					
	36	319					
lef-9	**37**	**503**	**40**	**507**	**62·9**	**ac62**	
	38	**141**	**41**	**147**	**51·4**	**ac68**	
NeleNPV *dr3*							2 × 29 bp
	39	*353*	*42*	*444*	*23·7*		
	40	*125*	*43*	*125*	*38·9*		ac75
	41	**76**	**44‡**	**82**	**58·4**	**ac76‡**	
vlf-1	**42**	**354**	**45**	**353**	**62·7**	**ac77**	
	43	**94**	**46**	**85**	**40·7**	**ac78**	
gp41	**44**	**270**	**47**	**312**	**62·4**	**ac80**	
	45	**175**	**48**	**176**	**72·7**	**ac81**	
p47	**46**	**389**	**49**	**386**	**59·7**	**ac40**	
NeseNPV *hr6*							11 × 45 bp
p74	**47**	**633**	**50**	**634**	**62·1**	**ac138**	
	48	*178*	*51*	*142*	*23·1*		DSRM
	49	*218*	*52*	*220*	*35·6*		C2H2 zinc-finger protein
pep	**50**	**407**	**53**	**386**	**45·7**	**ha57**	
	51	*192*	*54*	*184*	*48·1*		
pif-2	**52**	**384**	**55**	**384**	**74·3**	**ac22**	
	53	*104*	*56*	*115*	*62·9*		
lef-2	**54**	**195**	**57**	**200**	**35·7**	**ac6**	
lef-5	**55**	**235**	**58**	**230**	**58·0**	**ac99**	
38k	**56**	**285**	**59**	**304**	**57·7**	**ac98**	
p19	**57**	**168**	**60**	**170**	**61·5**	**ac96**	
helicase	**58**	**1134**	**61**	**1143**	**58·5**	**ac95**	
lef-4	**59**	**468**	**62**	**477**	**52·2**	**ac90**	
p49	**60**	**442**	**63**	**441**	**51·6**	**ac142**	
	61	*78*	*64*	*78*	*44·3*		
odv-e18	**62**	**85**	**65**	**83**	**84·5**	**ac143**	
odv-ec27	**63**	**262**	**66**	**260**	**72·4**	**ac144**	TMD
	64	**110**	**67**	**108**	**67·9**	**ac145**	
lef-1	**65**	**211**	**68**	**190**	**52·4**	**ac14**	
	66	**193**	**69**	**190**	**61·8**	**ac115**	
	67	**354**	**70**	**357**	**58·3**	**ac109**	
	68	*69*	*71*	*71*	*51·4*		SP or TMD
	69	*136*	*72*	*118*	*60·5*		RCC1
	70	*133*	*73*	*122*	*43·1*		RCC1

Table 2. cont.

Name	NeleNPV		NeseNPV			Baculovirus ORF*	Comments†
	ORF	Length (aa)	ORF	Length (aa)	Identity (%)		
			74	55			RCC1
	71	99					RCC1
	72	*283*	*75*	*273*	*33·9*		
			76	60			
	73	97					
	74	*265*	*77*	*284*	*36·5*		
	75	*209*	*78*	*143*	*29·2*		
pif	**76**	**530**	**79**	**519**	**62·5**	**ac119**	
NeleNPV *dr*4							3 × 43 bp
	77	**149**	**80**	**156**	**44·0**	**ac53**	
lef-8	**78**	**843**	**81**	**846**	**69·0**	**ac50**	
NeleNPV *dr*5							3 × 67 bp
NeleNPV *dr*6							2 × 45 bp
	79	64				xecn164	
	80	78					
			82	209			
	81	*148*	*83*	*145*	*67·8*		Densovirus capsid protein
vp91 capsid	**82**	**803**	**84**	**824**	**49·8**	**ac83**	
vp1054	**83**	**313**	**85**	**310**	**59·8**	**ac54**	
	84	*69*	*86*	*73*	*58·6*		
	85	*305*	*87*	*311*	*39·2*		
NeleNPV *dr*7							2 × 30 bp
NeleNPV *dr*8	86	86					2 × 160 bp
NeseNPV *dr*3							2 × 13 bp
	87	476					
NeleNPV *dr*9							2 × 43 bp
NeseNPV *dr*4			88	224			2 × 48 bp (in nese88)
vp39	**88**	**315**	**89**	**312**	**71·2**	**ac89**	
	89	*183*	*90*	*192*	*39·1*		Phosphotransferase

*AcMNPV homologous ORFs are listed preferentially.

†Possible homologues with significant BLASTP or VOCs matches, but with amino acid identity under 20 %, are indicated under 'Comments'.

‡nese44 is only an ac76 homologue by inference.

content and low similarity between ORFs in the NSRs. The G+C content of the NeleNPV and NeseNPV NSRs was also different, at 33·7 and 37·4 mol% G+C, respectively. Differences in nucleotide composition have been attributed to different levels of gene expression, differences in time of gene acquisition or codon usage of the host, a strategy to reduce the competition for nucleotides in viruses infecting the same host (Lange & Jehle, 2003), and insertion of DNA from a different origin (Desiere *et al.*, 2001). NeseNPV had an inversion relative to NeleNPV between nele22 and nele34 and a high degree of collinearity with NeleNPV for the remainder of the genome (Table 2; Fig. 1).

Repeat regions

A major difference between the two genomes was that NeseNPV contained baculoviral *hr*s as well as *dr*s, whereas NeleNPV contained only *dr*s. Most repeats were found within the NSRs, where NeseNPV had five *hr*s and two *dr*s and NeleNPV had one *dr* and three ORFs (nele2, nele7 and nele8) containing internal repeats (Table 2). The size difference between the two genomes may be partially attributed to the presence of six *hr*s in NeseNPV, accounting for 3669 bp. NeseNPV *hr* regions had a higher G+C content (49·8 mol%) than the overall genome (33·8 mol%), contributing to the slightly higher G+C content in NeseNPV compared with NeleNPV (33·3 %) and the higher G+C content of the NeseNPV NSR compared with the NeleNPV NSR. Without its *hr*s, the G+C content of NeseNPV was 33·0 mol%. The presence of repeats has been associated with major rearrangements, insertions and deletions in baculovirus genomes (Ahrens *et al.*, 1997). Even in closely related baculoviruses, such as *Rachiplusia ou* MNPV and AcMNPV or *Helicoverpa armigera* NPV isolates C1 and G4, differences are mainly

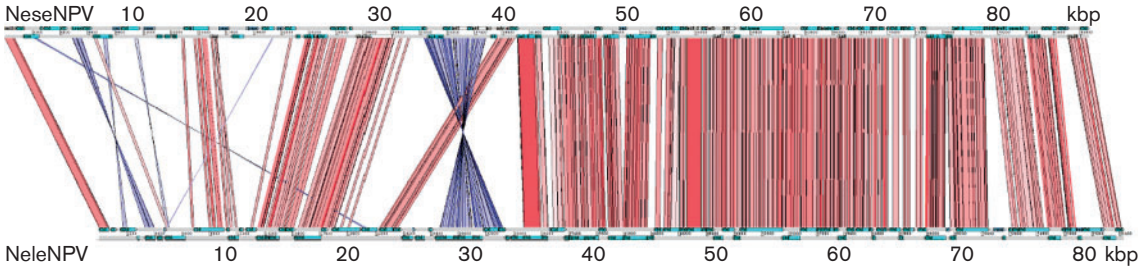

Fig. 1. Syntenic map. Comparison of the NeseNPV and NeleNPV genomes, indicating their overall level of collinearity. Purple lines indicate inversions, dark red indicates levels of identity among HSPs roughly above 40 % and pink below 30 %. Lines are between conserved regions, not necessarily ORFs. Genome position in kbp is shown above and below the map.

located near *hr* regions (Harrison & Bonning, 2003; Zhang *et al.*, 2005). The presence of multiple repeats in the NSRs, therefore, may account for some of the variation between the two genomes in this region.

iap genes

The close relationship between baculoviral *iap* genes and insect *iap* genes suggests that baculoviral *iap* genes may have been acquired through gene transfer from host insects (Huang *et al.*, 2000; Hughes & Friedman, 2003). NeseNPV and NeleNPV IAPs strengthen this hypothesis, as nese17 and nele11 showed top BLASTP matches to insect IAPs. They showed even closer matches to insect IAPs than to each other. nele11 contained two baculovirus IAP repeats (BIRs) and lacked a RING finger, whereas nese17 had one BIR and a zinc finger.

Membrane-fusion proteins

A striking feature of NeseNPV and NeleNPV is the lack of an identified membrane-fusion protein homologous to GP64 or to an F protein (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004). ld130 homologues, found in all group II NPVs and some group I NPVs, as well as in CuniNPV, may be the primordial baculovirus envelope-fusion proteins and GP64/67 homologues may be the functional replacements for ld130 homologues in group I NPVs (Rohrmann & Karplus, 2001). These proteins mediate the fusion of budded virus to cell membranes and the release of nucleocapsids (Pearson *et al.*, 2000, 2001). GP64 may also be required for the spread of infection from the insect gut to the haemocoel (Monsma *et al.*, 1996), a function perhaps not required in the hymenopteran viruses due to their restriction to the midgut. Membrane-fusion proteins generally have limited sequence similarity, making their identification difficult (Rohrmann & Karplus, 2001; Pearson & Rohrmann, 2002), but they usually contain a signal peptide, transmembrane domain, conserved cysteines and a furin-cleavage site (Kuzio *et al.*, 1999; Pearson *et al.*, 2000; Rohrmann & Karplus, 2001). There were no unidentified ORFs common to NeseNPV and NeleNPV that contained both signal peptides and transmembrane domains. nele18/nese26 and nele68/nese71 had transmembrane domains that overlapped potential signal peptides, but the ORFs were much shorter than ld130 or GP64 homologues and lacked conserved cysteines and a furin-cleavage site. Baculoviruses appear to be very adaptable and those lacking GP64 can utilize the envelope-fusion protein of *Vesicular stomatitis virus* as their fusion protein, suggesting that the ability of baculoviruses to enter cells could be accommodated by a variety of envelope proteins (Mangor *et al.*, 2001; Pearson & Rohrmann, 2002). It is therefore possible that another hymenopteran baculovirus ORF(s) may act as a functional replacement for a membrane-fusion protein if one is required by hymenopteran baculoviruses.

Shared ORFs between NeseNPV and NeleNPV

Several ORFs not previously identified in lepidopteran or dipteran baculoviruses were reported independently in NeseNPV and NeleNPV and included a trypsin-like serine protease (nele6/nese7, 72·6 % amino acid identity), a zinc finger-like protein (nele49/nese52, 35·6 % amino acid identity), three proteins homologous to regulators of chromosome condensation proteins (RCC1) in NeleNPV and two in NeseNPV (nele69/nese72, 60·5 % amino acid identity; nele70/nese73, 43·1 % amino acid identity; nele71), a densovirus-like capsid protein (nele81/nese83, 67·8 % amino acid identity) and a phosphotransferase homologue (nele89/nese90, 39·1 % amino acid identity). Some of these ORFs may be specific to hymenopteran baculoviruses and provide them with a selective advantage.

Trypsin-like serine protease. NeseNPV and NeleNPV are the first reported baculoviruses with trypsin-like serine proteases. Similar proteins have been identified in actinomycetes and bacteria and in many eukaryotes, including insects (Ross *et al.*, 2003). nele6 contained the trypsin catalytic triad, histidine, aspartic acid and serine, as well as the six conserved cysteines reported in nese7 (Garcia-Maruniak *et al.*, 2004). Both ORFs shared top BLASTP matches to similar proteins from insects. Hymenopteran trypsin-like serine proteases with BLASTP matches to nele6/nese7 included *A.*

mellifera proteins (GenBank accession nos XP_394076 and XP_397087). The honeybee protein GenBank XP_394076 shared 46·2 % amino acid identity with nele6 and 47·1 % with nese7, values higher than the amino acid identities of most hymenopteran baculovirus ORFs to lepidopteran NPV homologues. All three phylogenetic analyses (maximum likelihood with PHYML, TREE-PUZZLE and maximum parsimony) grouped the hymenopteran baculovirus trypsins with the *A. mellifera* protein GenBank XP_394076 and showed relatedness to other insect trypsin-like proteins. The best-resolved consensus was obtained with the non-parametric bootstrap tree with PHYML using the Jones–Taylor–Thornton (JTT) substitution model for amino acids (Jones *et al.*, 1992) and is shown in Fig. 2. Bootstrap values >50 % obtained with the other two methods are also shown in Fig. 2. All three methods supported the idea that the trypsin genes of the hymenopteran baculoviruses were acquired by an ancestral *Neodiprion* baculovirus via horizontal transfer, possibly from a host. The G+C contents of nele6 (43·2 mol%) and nese7 (43·5 mol%) were much higher than the overall G+C content of either genome. Differences in the G+C content of an ORF relative to a genome could be an indication of insertion of DNA from a different origin (Desiere *et al.*, 2001). There is evidence that several baculovirus genes have been transferred horizontally from eukaryotes, possibly from their hosts or bacterial sources (Hughes & Friedman, 2003). Some genes originally derived from a host may have different functions in the virus due to viral adaptation, as was found recently for the baculovirus F protein (Lung & Blissard, 2005).

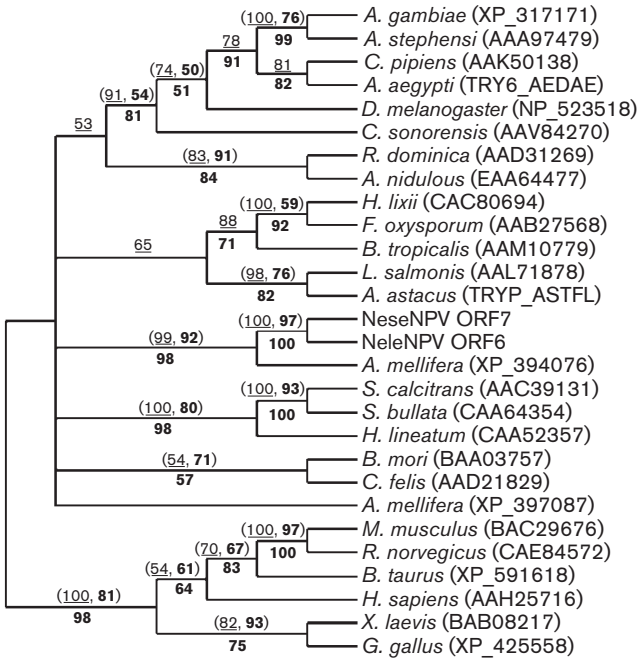

Fig. 2. Consensus tree for trypsin-like serine proteases. Phylogenetic tree showing the percentage level of support obtained after 1000 replications of non-parametric bootstrap (underlined above branches) using PHYML (global tree search by maximum likelihood). The levels of support with the other two methods are also shown: (i) the quartet-method search with TREE-PUZZLE (bold above branches) and (ii) global parsimony with PAUP and 1000 bootstrap replicates (below branches). Only nodes with support above 50 % are annotated.

In order to determine whether nele6/nese7 could be functional, their 3D trypsin models were overlapped structurally with a bovine β-trypsin (PDB file 2PTC) (Marquart *et al.*, 1983) to compare the overall conformation of the molecules and the structure of their active sites. A Cα overlap of nese7 and β-trypsin gave a square root of mean square deviation (RMSD) of 2·3 Å (Fig. 3a). A comparison of a Cα overlap of nese7 and nele6 gave an RMSD of 1·3 Å (Fig. 3b). The structural-homology modelling of NeseNPV and NeleNPV trypsins showed that their conformations were similar to that of the bovine β-trypsin. Fig. 3(c) shows that the predicted catalytic residues of nese7 (His70, Asp117 and Ser213) and nele6 (His71, Asp118 and Ser214) overlapped structurally with the catalytic triad of β-trypsin, suggesting that the viral trypsins may be functional proteins. Experimental assays are under way to ascertain functionality.

Zinc-finger protein. Zinc-finger domains are nucleic acid-binding structures, first identified in the *Xenopus* transcription factor TFIIIA, and are one of the most diverse superfamilies of nucleic acid-binding proteins in eukaryotes (Böhm *et al.*, 1997). C2H2 zinc fingers are composed of 25–30 amino acid residues, including two conserved cysteines and two conserved histidine residues in a C–2–C–12–H–3–H motif that binds a zinc ion (Knight & Shimeld, 2001). nele49 and nese52 contained four C2H2 zinc-finger domains between aa 2 and 22, 26 and 49, 54 and 75, and 81 and 101 in nele49 and aa 2 and 22, 28 and 51, 56 and 77, and 83 and 106 in nese52. They shared an overall amino acid identity of 35·6 %, but identity over the zinc-finger regions was much higher at 50 %. It has been found that a tandem array comprising a minimum of two zinc fingers is required for sequence-specific, high-affinity DNA binding (Böhm *et al.*, 1997). Inference of evolutionary history is difficult for zinc-finger proteins, due to the conservation of key residues critical for the structure of the domain and the repetition of the C2H2 motif in individual genes. By using percentage amino acid identity over zinc-finger regions, it has been suggested that a similarity score >45 % indicates a relationship not due to baseline identity and that all sequences showing an identity score >55 % are in orthology groups (Knight & Shimeld, 2001). It is therefore likely that nese52 and nele49 are related. No significant virus matches were found for nele49 or nese52; instead, top matches included a zinc-finger protein from *Drosophila melanogaster* (GenBank accession no. NP_609448; 36 % amino acid identity, 40/109 aa) for nele49 and an *Anopheles gambiae* zinc-finger protein (GenBank accession no. EAA00343; 36 % amino acid identity, 38/104 aa) for nese52. They also

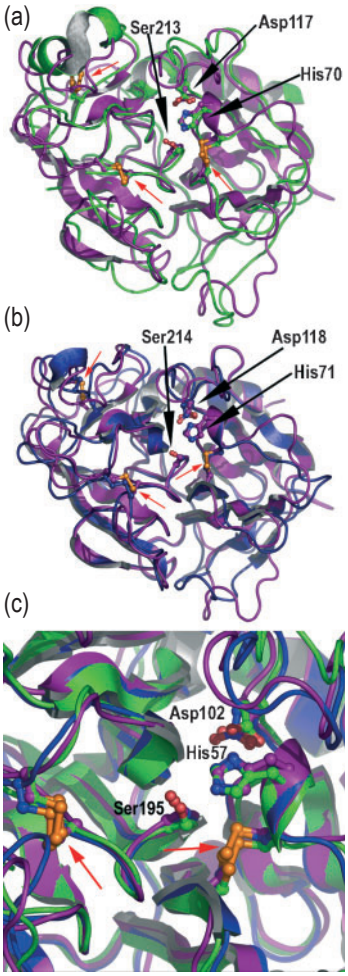

Fig. 3. Three-dimensional protein structure of trypsin-like serine proteases. Ribbon models of NeseNPV ORF7 (purple) and NeleNPV ORF6 (blue) predicted protein structures compared with bovine β-trypsin (green) (PDB file 2PTC). Catalytic residues and disulfide bonds are drawn as ball-and-stick models. Disulfide bonds are highlighted with red arrows. (a) Cα overlap of bovine β-trypsin and nese7. Catalytic residues histidine (His), aspartic acid (Asp) and serine (Ser) for nese7 are labelled. (b) Cα overlap of nese7 and nele6. Catalytic residues for nele6 are labelled. (c) Enlargement of the nese7, nele6 and bovine β-trypsin active-site triad. Catalytic residues for β-trypsin are labelled.

showed significant BLASTP matches to several hymenopteran zinc-finger proteins, including *A. mellifera* proteins (GenBank accession nos XP_395651 and XP_39083), with seven and 15 zinc fingers, respectively. An alignment of individual zinc fingers from the above sequences is shown (Fig. 4). The combined evidence of sequence identity between nele49 and nese52, the conserved number and position of zinc fingers and the similar location of both ORFs within the genomes suggest that these proteins may play a hitherto unknown role in the hymenopteran baculoviruses.

```
                                        *          20
NeleNPV ORF49/2-22                  :  PTCEICHLNVVSR--FKRHHQQH
NeseNPV ORF52/2-22                  :  PFCGICETSIKSR--FKRHMECH
D. melanogaster NP_609448/254-276   :  FQCTHCEASFPNAGDLSKHVRSH
A. gambiae EAA00343/52-74           :  YQCEVCQKDFMGTNDLRKHLRIH
A. mellifera XP_395651/203-225      :  YKCNICNKTFTFQQSYHKHRLYH
A. mellifera XP_393083/92-114       :  IQCSLCKKWFLNNDSMVTHLRMH
                                       C  C           H    H
```

Fig. 4. Alignment of zinc-finger domains. The first zinc-finger domain in NeleNPV ORF49 and NeseNPV ORF52, aligned with representative zinc-finger domains from insect proteins. Black shading represents 100 % amino acid identity, medium grey, 80 % and light grey, 60 %. Names of organisms, GenBank accession numbers and sequence locations are shown on the left and the amino acid consensus underneath.

Regulator of chromosome condensation proteins. Regulator of chromosome condensation proteins (RCC1) are DNA-binding or chromatin-associated proteins that play a role in regulating the onset of chromosome condensation at the level of transcription or mRNA maturation (Ohtsubo *et al.*, 1989). They bind to chromatin and interact with ran, a nuclear GTP-binding protein, to promote the loss of bound GDP and the uptake of fresh GTP, thus acting as a guanine nucleotide-dissociation stimulator. The interaction of RCC1 with ran probably plays an important role in the regulation of gene expression. In addition to being essential for DNA replication, the initiation of mitosis and the transition from mitosis to G1, the RCC1/ran proteins are also involved in protein import into the nucleus (Huang *et al.*, 1997).

NeseNPV and NeleNPV have three RCC1 homologues located as sequential ORFs in their genomes and are the only viruses with RCC1/BLIP domains noted in the InterPro taxonomic coverage for RCC1 proteins, although several insect groups have these proteins (InterPro IPR000408). nele69/nese72 shared top BLASTP matches with several insect RCC1 proteins, including *D. melanogaster* BJ1 (GenBank accession no. P25171; nele69, 35·8 %; nese72, 40·3 % amino acid identity). The *D. melanogaster* RCC1 protein, BJ1, has been found to be functionally equivalent to the vertebrate RCC1 proteins (Ohtsubo *et al.*, 1991). The top BLASTP match for nele70/nese73 was the RCC1 protein from *A. gambiae* (GenBank accession no. XP_310273; nele70, 28·4 %; nese73, 33·0 % amino acid identity). Insect RCC1 matches were also found with nele71 and nese74, but these ORFs shared low amino acid identity with each other (10·7 %) and were not considered homologues to each other, but both were considered RCC1 homologues. nele71 was previously identified as an RCC1 homologue (Lauzon *et al.*, 2004), but nese74 was not (Garcia-Maruniak *et al.*, 2004). By using TBLASTX, the nucleotide sequence from nele69 to nele71 and from nese72 to nese74 showed top match to an *A. mellifera* RCC1 protein (GenBank accession no. XM_394158).

RCC1 proteins contain seven tandem repeats of a domain composed of 50–60 aa. The repeats make up the major part

of the length of the protein (Ohtsubo *et al.*, 1989). Only one copy of the RCC1 repeat domains was found in each of the NeseNPV and NeleNPV RCC1 ORFs, making the ORFs smaller (55–136 aa) than other RCC1 proteins (*D. melanogaster* BJ1, 547 aa). The total size of the three potential RCC1 proteins in NeseNPV (295 aa) and NeleNPV (368 aa) was still smaller than other RCC1 proteins. Two signature patterns are present in RCC1 proteins. The first and most conserved is found in the N terminus of the second repeat and has the pattern G-x-N-D-x(2)-(AV)-L-G-R-x-T (PROSITE PS00625; Hulo *et al.*, 2004). nele69 contained a perfect match to this consensus pattern and nese52 had a close, but imperfect match (Fig. 5). The second consensus pattern is derived from conserved positions in the C-terminal part of each repeat. Potential matches to the second consensus pattern were not as close in the NeseNPV and NeleNPV RCC1 homologues. The sequence of the repeated domain of RCC1 proteins appears to be well-conserved through evolution (Ohtsubo *et al.*, 1991). The presence of RCC1-like genes in various insects and in the hymenopteran baculoviruses suggests the possibility of a horizontal transfer and a possible role in the biology of hymenopterans. The presence of three small ORFs each containing only one RCC1 repeat, however, might also mean that a mutation has occurred, resulting in a frameshift making the RCC1 homologues non-functional.

Densovirus capsid protein. Densovirus genomes, which contain a single molecule of linear, negative-sense or positive-sense, single-stranded DNA, are about 5–6 kb in length and have long, inverted terminal repeats (Büchen-Osmond, 2003). Both nele81 and nese83 showed strong BLASTP matches to densovirus structural proteins. Top BLASTP matches were to viral protein 1-4 from *Casphalia extranea* densovirus (GenBank accession no. NP_694840) (nele81, 30·2 %; nese83, 31·5 % amino acid identity) and a capsid protein from *Bombyx mori* densovirus (GenBank accession no. NP_694837) (nele81, 24·8 %; nese83, 29·5 % amino acid identity). nele81 and nese83, however, were significantly shorter than these densovirus proteins. Most densoviruses infect a variety of larval tissues, but not the midgut. Interestingly, densoviruses from *B. mori* and *C. extranea* are an exception and are found predominantly in the columnar cells of midgut epithelium (Fédière *et al.*, 2002). The functionality of the densovirus homologues in the hymenopteran baculoviruses is not known, but work is presently under way to investigate this.

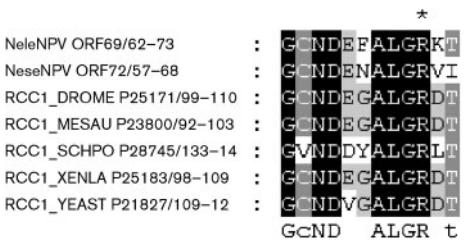

Fig. 5. Alignment of first RCC1 signature patterns. The regulator of condensation signature 1 pattern, G-x-N-D-x(2)-[AV]-L-G-R-x-T, from NeleNPV ORF 69 and NeseNPV ORF 72 were compared with those from other RCC1 proteins. Black shading represents 100 % amino acid identity, medium grey, 80 % and light grey, 60 %. Names of organisms, GenBank accession numbers and sequence locations are shown on the left and the amino acid consensus underneath.

Phosphotransferase. BLAST analysis showed that nele89 and nese90 had homology with the RNA 2′-phosphotransferase KptA/TPT family, a group of proteins involved in tRNA splicing. Phosphotransferases have not been reported previously in viruses, but are found in eukaryotes and a limited number of eubacteria and archaeal organisms. Eubacteria are not known to splice tRNA, suggesting an unknown function for this protein family (Spinelli *et al.*, 1998, 1999). Phylogenetic analysis shows no evidence for a recent horizontal transfer of the phosphotransferase into eubacteria, but suggests that it has been present in this group since close to the time when the eukaryotes, eubacteria and archaea diverged (Spinelli *et al.*, 1998). A most-parsimonious tree including nele89 and nese90 confirmed this hypothesis and showed that nele89 and nese90 group in the eukaryote clade (data not shown). After nese90, the top BLASTP match for nele89 was a phosphotransferase protein from *D. melanogaster* (GenBank accession no. NP_788477). This protein showed 36·7 % amino acid identity to nele89 and 32·7 % to nese90. Only one insect phosphotransferase was found, so it is not known whether a horizontal transfer could have occurred from an insect host or perhaps another eukaryote or an intermediary. Sequence blocks conserved in phosphotransferase proteins (Spinelli *et al.*, 1998) were found in both nele89 and nese90, suggesting that they may share a common structure with known phosphotransferase proteins and might therefore be functional (Fig. 6).

Other ORFs found only in NeseNPV and NeleNPV

The shared ORFs already mentioned accounted for only a few of the ORFs found only in NeseNPV and NeleNPV. Analysis of the remaining ORFs indicated that some may have baculovirus identity too low for clear identification and some may be specific to hymenopteran baculoviruses. Several ORFs were reported previously as potential baculovirus homologues, but amino acid identities were too low for them to be accepted as clearly identified homologues. These included nese26/*odv-e25* (ac94), nese29/*desmoplakin* (ac66), nese43/ac75 and nele53/*p10* (ac137) (Garcia-Maruniak *et al.*, 2004; Lauzon *et al.*, 2004). Analysis with the VOCs program did not support nele18/nese26 (41·9 % amino acid identity) as being *odv-e*25 homologues, although BLASTP searches showed low-overlapping matches to envelope glycoproteins. nele53/nese56 (62·9 % amino acid identity) also did not appear as *p10* homologues by using VOCs, although they shared features common to P10 proteins,

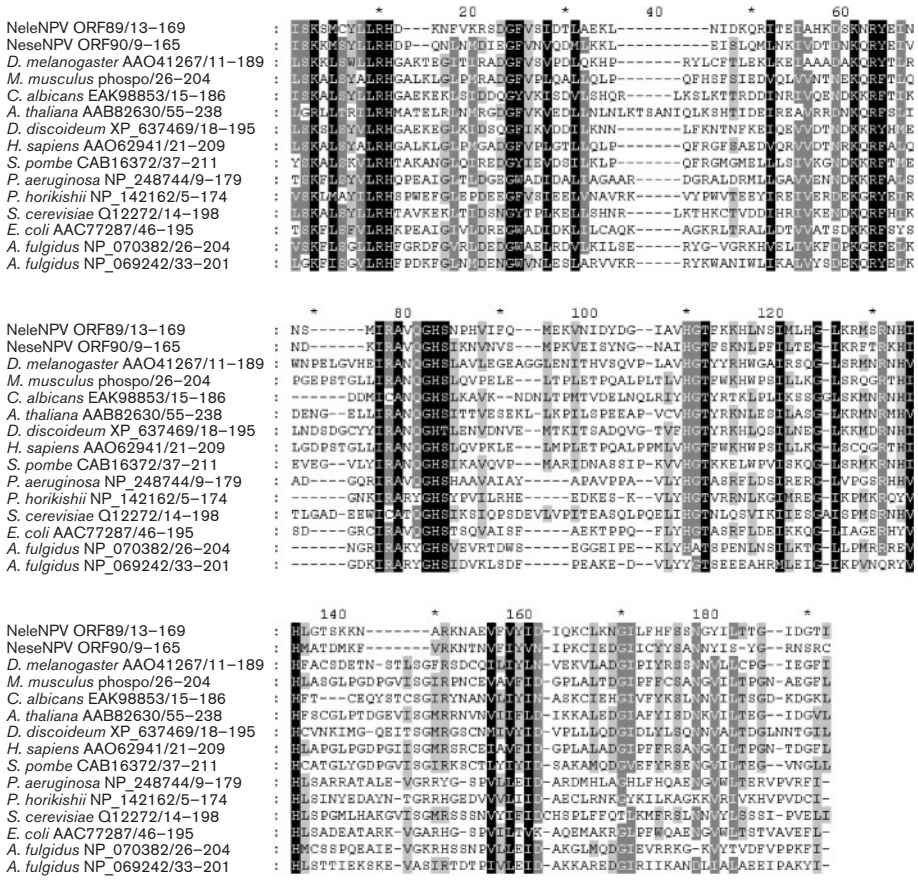

Fig. 6. Amino acid alignment of phosphotransferase homologues. The PTS-2RNA domains of phosphotransferase homologues were aligned with MEGALIGN CLUSTAL W, using default conditions. Black shading indicates 100 % equivalent residues (i.e. conservative substitutions), medium grey, 80 % and light grey, 60 %. Proteins include the top BLASTP matches for NeleNPV ORF 89 and NeseNPV ORF 90, plus proteins used in Fig. 3 of the paper by Spinelli *et al.* (1998). Names of organisms, GenBank accession numbers and locations of sequence are shown on the left.

including a late promoter, correct size range, a conserved A at position −3, location near p74 and helix-forming residues. The VOCs program indicated that nele21/nese29 may be desmoplakin homologues and nele40/nese43 may be ac75 homologues, but amino acid identities were < 20 %. These ORFs were in the expected genomic location, so the possibility remains of them being baculovirus homologues. VOCs analysis suggested that nele9/nese16 were desmoplakin homologues, that nele35 was a repeated ORF similar to a BRO protein and that nele61/nese64 were *Phthorimaea operculella* GV ORF 117 homologues. These potential matches are listed under comments in Table 2, but were not considered as being clearly identified. Some of the remaining shared ORFs had high levels of amino acid identity (nele53/nese56, 62·9 %) or shared similar features (nele48/nese51, double-stranded RNA-binding motifs), implying that they may be important in the biology of hymenopteran baculoviruses.

In conclusion, NeseNPV and NeleNPV appeared to be related much more closely to each other than to the lepidopteran baculoviruses or to CuniNPV. Both were small in size, were AT-rich, had fewer ORFs than other baculoviruses and their arrangement of ORFs was basically collinear. Neither had an identified membrane-fusion protein, previously considered a core baculovirus protein, and both shared many ORFs not found previously in lepidopteran or dipteran baculoviruses. A few of the unknown ORFs found only in NeleNPV and NeseNPV, such as the trypsin-like serine protease homologues, might have been transferred horizontally from insect hosts. NeseNPV and NeleNPV had NSRs between their *polyhedrin* and *dbp* genes that contained a large number of repeats and unique ORFs. The viruses differed from each other in that ORFs present in their NSRs were variable and their NSRs had different G+C contents. Across their genomes, NeleNPV contained 20 ORFs not found in NeseNPV and NeseNPV had 21 ORFs not present in NeleNPV. NeleNPV also had an inversion relative to NeseNPV between nele23 and nele34 and NeseNPV had *dr*s and *hr*s, whereas NeleNPV had only *dr*s. Although NeseNPV and NeleNPV were related more closely to each other than to other baculoviruses, their

mean amino acid identity was not high, suggesting that NeseNPV and NeleNPV may be related more closely to other hymenopteran baculoviruses than to each other. Based on the comparison of the polyhedrin sequences from *Neodiprion abietis* NPV (NeabNPV), another New World sawfly baculovirus, NeabNPV appeared to be related more closely to NeleNPV than to NeseNPV. Perhaps some of the differences between NeleNPV and NeseNPV might be due to *N. lecontei* being a New World species and *N. sertifer* an Old World pest. The two viruses would have had little geographical overlap until the introduction of NeseNPV to North America. The genomic sequence of another hymenopteran baculovirus will help to confirm which ORFs may be unique to hymenopteran baculoviruses, provide further information on the evolution of hymenopteran baculoviruses and provide further support for a separate grouping of hymenopteran baculoviruses in the family *Baculoviridae*.

ACKNOWLEDGEMENTS

This research was partly supported by grants from Genome Canada through the Ontario Genomics Institute, the Canadian Biotechnology Strategy Fund and the United States Department of Agriculture grant 2001-35302-11006. We would like to thank Drs Jenny Cory and Yoshifumi Hashimoto for their valuable comments.

REFERENCES

Afonso, C. L., Tulman, E. R., Lu, Z., Balinsky, C. A., Moser, B. A., Becnel, J. J., Rock, D. L. & Kutish, G. F. (2001). Genome sequence of a baculovirus pathogenic for *Culex nigripalpus*. *J Virol* **75**, 11157–11165.

Ahrens, C. H., Russell, R. 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.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Bird, F. T. (1953). The use of a virus disease in the biological control of the European pine sawfly, *Neodiprion sertifer* (Geoffr.). *Can Entomol* **85**, 437–446.

Bird, F. T. (1961). Transmission of some insect viruses with particular reference to ovarial transmission and its importance in the development of epizootics. *J Insect Pathol* **3**, 352–380.

Böhm, S., Frishman, D. & Mewes, H. W. (1997). Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. *Nucleic Acids Res* **25**, 2464–2469.

Brown, D. A. (1982). Two naturally occurring nuclear polyhedrosis virus variants of *Neodiprion sertifer* Geoffr. (Hymenoptera; Diprionidae). *Appl Environ Microbiol* **43**, 65–69.

Büchen-Osmond, C. (editor) (2003). *Densovirus*. In *ICTVdB – The Universal Virus Database*, version 3, chapter 00.050.2.01. ICTVdB Management, The Earth Institute and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA. http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdb/index.htm

Cunningham, J. C., DeGroot, P. & McPhee, J. R. (1984). Lecontvirus: a viral insecticide for control of redheaded pine sawfly, *Neodiprion lecontei*. In *Biological Control Methods* (Technical Note No. 2). Sault Ste Marie, ON: Canadian Forest Service.

De Groot, P. & Cunningham, J. C. (1983). *Aerial Spray Trials with a Baculovirus to Control Red Headed Pine Sawfly in Ontario in 1979 and 1980* (Canadian Forest Service Information Report FPM-X-63). Sault Ste Marie, ON: Canadian Forest Service.

DeLano, W. L. (2002). The PyMOL molecular graphics system. San Carlos, CA: DeLano Scientific. http://pymol.sourceforge.net/

Desiere, F., Mahanivong, C., Hillier, A. J., Chandry, P. S., Davidson, B. E. & Brüssow, H. (2001). Comparative genomics of lactococcal phages: insight from the complete genome sequence of *Lactococcus lactis* phage BK5-T. *Virology* **283**, 240–252.

Federici, B. A. (1997). Baculovirus pathogenesis. In *The Baculoviruses*, pp. 33–60. Edited by L. K. Miller. New York: Plenum.

Fédière, G., Li, Y., Zádori, Z., Szelei, J. & Tijssen, P. (2002). Genome organization of *Casphalia extranea* densovirus, a new *Interavirus*. *Virology* **292**, 299–308.

Garcia-Maruniak, A., Maruniak, J. E., Zanotto, P. M. A., 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.

Guindon, S. & Gascuel, O. (2003). A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.

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.

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.

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.

Herniou, E., Olszewski, J. A., O'Reilly, D. R. & Cory, J. S. (2004). Ancient coevolution of baculoviruses and their insect hosts. *J Virol* **78**, 3244–3251.

Huang, S., Mayeda, A., Krainer, A. R. & Spector, D. L. (1997). RCC1 and nuclear organization. *Mol Biol Cell* **8**, 1143–1157.

Huang, Q., Deveraux, Q. L., Maeda, S., Salvesen, G. S., Stennicke, H. R., Hammock, B. D. & Reed, J. C. (2000). Evolutionary conservation of apoptosis mechanisms: lepidopteran and baculoviral inhibitor of apoptosis proteins are inhibitors of mammalian caspase-9. *Proc Natl Acad Sci U S A* **97**, 1427–1432.

Huber, J. (1986). Use of baculoviruses in pest management programs. In *The Biology of Baculoviruses*, vol. II, *Practical Application for Insect Control*, pp. 182–197. Edited by R. R. Grandados & B. A. Federici. Boca Raton, FL: CRC Press.

Hughes, A. L. & Friedman, R. (2003). Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. *Mol Biol Evol* **20**, 979–987.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**, 275–282.

Knight, R. D. & Shimeld, S. M. (2001). Identification of conserved C2H2 zinc-finger gene families in the Bilateria. *Genome Biol* **2**, research0016.1–0016.8. doi:10.1186/gb-2001-2-5-research0016

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.

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

Lung, O. & Blissard, G. W. (2005). A cellular *Drosophila melanogaster* protein with similarity to baculovirus F envelope fusion proteins. *J Virol* **79**, 7979–7989.

Mangor, J. T., Monsma, S. A., Johnson, M. C. & Blissard, G. W. (2001). A GP64-null baculovirus pseudotyped with vesicular stomatitis virus G protein. *J Virol* **75**, 2544–2556.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr Sect B Struct Sci* **39**, 480–490.

Monsma, S. A., Oomens, A. G. P. & 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.

Ohtsubo, M., Okazaki, H. & Nishimoto, T. (1989). The RCC1 protein, a regulator for the onset of chromosome condensation locates in the nucleus and binds to DNA. *J Cell Biol* **109**, 1389–1397.

Ohtsubo, M., Yoshida, T., Seino, H., Nishitani, H., Clark, K. L., Sprague, G. F., Jr, Frasch, M. & Nishimoto, T. (1991). Mutation of the hamster cell cycle gene *RCC1* is complemented by the homologous genes of *Drosophila* and *S. cerevisiae*. *EMBO J* **10**, 1265–1273.

Pearson, M. N. & Rohrmann, G. F. (2002). Transfer, incorporation, and substitution of envelope fusion proteins among members of the *Baculoviridae*, *Orthomyxoviridae*, and *Metaviridae* (insect retrovirus) families. *J Virol* **76**, 5301–5304.

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.

Pearson, M. N., Russell, R. L. Q. & Rohrmann, G. F. (2001). Characterization of a baculovirus-encoded protein that is associated with infected-cell membranes and budded virions. *Virology* **291**, 22–31.

Purcell, W. P. & Singer, J. A. (1967). A brief review and table of semiempirical parameters used in the Hückel molecular orbital method. *J Chem Eng Data* **12**, 235–246.

Rohrmann, G. F. & Karplus, P. A. (2001). Relatedness of baculovirus and *gypsy* retrotransposon envelope proteins. *BMC Evol Biol* **1**, 1.

Ross, J., Jiang, H., Kanost, M. R. & Wang, Y. (2003). Serine proteases and their homologs in the *Drosophila melanogaster* genome: an initial analysis of sequence conservation and phylogenetic relationships. *Gene* **304**, 117–131.

Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**, 5857–5864.

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res* **28**, 231–234.

Spinelli, S. L., Malik, H. S., Consaul, S. A. & Phizicky, E. M. (1998). A functional homolog of a yeast tRNA splicing enzyme is conserved in higher eukaryotes and in *Escherichia coli*. *Proc Natl Acad Sci U S A* **95**, 14136–14141.

Spinelli, S. L., Kierzek, R., Turner, D. H. & Phizicky, E. M. (1999). Transient ADP-ribosylation of a 2′-phosphate implicated in its removal from ligated tRNA during splicing in yeast. *J Biol Chem* **274**, 2637–2644.

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), version 4. Sunderland, MA: Sinauer Associates.

Volkman, L. E., Blissard, G. W., Friesen, P., Keddie, B. A., Possee, R. & Theilmann, D. A. (1995). Family *Baculoviridae*. In *Virus Taxonomy: Sixth Report of the International Committee on Taxonomy of Viruses*, pp. 104–113. Edited by F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo & M. D. Summers. New York: Springer.

Zhang, C.-X., Ma, X.-C. & Guo, Z.-J. (2005). Comparison of the complete genome sequence between C1 and G4 isolates of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus. *Virology* **333**, 190–199.