Journal of General Virology (2014), **95**, 2297–2309 DOI 10.1099/vir.0.064766-0

Genomic diversity in European *Spodoptera exigua* multiple nucleopolyhedrovirus isolates

Julien Thézé,[1]† Oihana Cabodevilla,[2]† Leopoldo Palma,[2] Trevor Williams,[3] Primitivo Caballero[2,4] and Elisabeth A. Herniou[1]

Correspondence
Elisabeth A. Herniou
elisabeth.herniou@univ-tours.fr

[1]Institut de Recherche sur la Biologie de l'Insecte, CNRS UMR 7261, Université François-Rabelais de Tours, UFR Sciences et Techniques, 37200 Tours, France

[2]Instituto de Agrobiotecnología, CSIC-Gobierno de Navarra, 31192 Mutilva Baja, Navarra, Spain

[3]Instituto de Ecología AC, Xalapa, Veracruz 91070, Mexico

[4]Departamento de Producción Agraria, Universidad Pública de Navarra, 31006 Pamplona, Navarra, Spain

Received 18 February 2014
Accepted 19 May 2014

Key virus traits such as virulence and transmission strategies rely on genetic variation that results in functional changes in the interactions between hosts and viruses. Here, comparative genomic analyses of seven isolates of *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV) with differing phenotypes were employed to pinpoint candidate genes that may be involved in host–virus interactions. These isolates obtained after vertical or horizontal transmission of infection in insects differed in virulence. Apart from one genome containing a *piggyBac* transposon, all European SeMNPV isolates had a similar genome size and content. Complete genome analyses of single nucleotide polymorphisms and insertions/deletions identified mutations in 48 ORFs that could result in functional changes. Among these, 13 ORFs could be correlated with particular phenotypic characteristics of SeMNPV isolates. Mutations were found in all gene functional classes and most of the changes we highlighted could potentially be associated with differences in transmission. The regulation of DNA replication (helicase, lef-7) and transcription (lef-9, p47) might be important for the establishment of sublethal infection prior to and following vertical transmission. Virus–host cell interactions also appear instrumental in the modulation of viral transmission as significant mutations were detected in virion proteins involved in primary (AC150) or secondary infections (ME35) and in apoptosis inhibition (IAP2, AC134). Baculovirus populations naturally harbour high genomic variation located in genes involved at different levels of the complex interactions between virus and host during the course of an infection. The comparative analyses performed here suggest that the differences in baculovirus virulence and transmission phenotypes involve multiple molecular pathways.

INTRODUCTION

What explains the variation in transmissibility of a virus in a particular host species? A genetic basis to transmissibility is expected, but is usually elusive to define in the absence of a formal description of gene–host interactions (Bangham *et al.*, 2008; Wilfert & Jiggins, 2010). Studying the genomes of phenotypically variable pathogen strains can prove effective for pinpointing genes that are important in host–pathogen interactions at the population level (Allen & Little, 2009; Costa *et al.*, 2009).

Baculoviruses are large dsDNA viruses that are occluded into proteinaceous occlusion bodies (OBs) for horizontal transmission. OBs are pathogenic to certain species of insects, especially those in the order Lepidoptera (Gröner, 1986; Martignoni & Iwai, 1981). These viruses form the basis for a number of biological insecticides employed in the control of caterpillar pests of forests and field crops (Moscardi, 1999). The effectiveness of biological insecticides and the prevalence of disease in natural insect populations depend in large part on the variation in traits relevant to horizontal transmission, such as OB pathogenicity and OB productivity, associated with particular genotypes (Cory & Myers, 2003; Erlandson, 2009).

Spodoptera exigua multiple nucleopolyhedrovirus (SeMNPV) is a virus species from the group II viruses of the genus *Alphabaculovirus* (Herniou *et al.*, 2011). Productive infection by SeMNPV is specific to the beet armyworm *Spodoptera exigua* (Simón *et al.*, 2004), and this virus forms

†These authors contributed equally to this work.

The GenBank/EMBL/DDBJ accession numbers for the seven SeMNPV isolates are HG425343–HG425349.

the active ingredient of biological insecticides sold in the USA, Europe and Thailand (Bianchi *et al.*, 2001; Kolodny-Hirsch *et al.*, 1997; Smits *et al.*, 1987).

A Floridian strain of SeMNPV comprises defective genotypes that reduce the insecticidal properties of the isolate (Muñoz *et al.*, 1998). Populations of *S. exigua* from Almería in southern Spain are significantly more susceptible to a native Almerian SeMNPV strain (SeMNPV-SP2) compared with the Floridian strain (Belda *et al.*, 2000). SeMNPV-SP2 was subjected to field testing (Lasa *et al.*, 2007) and a product containing a mixture of genotypes has now been commercialized under the name Virex (Biocolor). Rapid uptake of virus-based insecticides by greenhouse growers and natural SeMNPV infections in *S. exigua* populations mean that SeMNPV populations are now present in most greenhouses in Almería. Moreover, infections are observed in the absence of application of these products because, like other baculoviruses, SeMNPV infections can be transmitted vertically from parent to offspring, and because OBs persist in greenhouse soil substrates and can be transferred onto crops and consumed by susceptible caterpillars (Murillo *et al.*, 2007). This unique field situation is the theatre for a set of continuous host–pathogen interactions.

High levels of genotypic heterogeneity have been observed in natural SeMNPV populations that originated from groups of virus-killed larvae collected in the field (Muñoz *et al.*, 1999), from greenhouse soils (Murillo *et al.*, 2007) or from the progeny of field-collected adult moths that harboured unapparent infections (Cabodevilla *et al.*, 2011). Four different genotypes (SP2-A–SP2-D) were obtained by cloning from the Almerian field isolate SeMNPV-SP2 (Muñoz *et al.*, 1999). SP2-A was the predominant variant in the SP2 strain, used as a bioinsectide in greenhouses (Muñoz *et al.*, 1999). Nine different genotypes (Se-G24–Se-G29 and Se-M1–Se-M3), presumably derived from SP2, were subsequently identified from greenhouse soil samples (Murillo *et al.*, 2007). As OBs in the soil can only achieve transmission if they are transported back onto plant surfaces and then eaten by susceptible insects, these genotypes were assumed to result from horizontal transmissions (Cabodevilla *et al.*, 2011). Additional isolates were obtained from the progeny of field-collected moths that died from virus infection in the laboratory (SeAl1 and SeAl2) or were isolated from a persistently infected laboratory colony (SeOx4–SeOx6). These infections could best be explained by vertical transmission of the virus from parents to their offspring (Cabodevilla *et al.*, 2011).

Selected biological traits of six of these isolates [VT-SeAl1, VT-SeAl2, VT-SeOx4 (obtained after vertical transmission, prefix VT) and HT-SeG24, HT-SeG25 and HT-SeG26 (obtained after horizontal transmission, prefix HT)] were compared. However, only HT-SeG25 was shown to be unable to sustain covert infection (Cabodevilla *et al.*, 2011). The isolates HT-SeG25 and HT-SeG26 were the most pathogenic (lower lethal concentration metrics) isolates. Isolate VT-SeOx4 had the fastest speed of kill compared with other isolates, whereas VT-SeAl1 was the slowest. In terms of yield, measured as OB production per insect, VT-SeAl2 was the most productive isolate (Cabodevilla *et al.*, 2011).

In order to understand the links between the genetic variation observed in SeMNPV isolates and phenotypic traits, such as pathogenicity, speed of kill and yield, the genomes of seven biologically distinct European SeMNPV isolates were sequenced and compared to identify potential associations between variations in phenotypic traits and vertical or horizontal transmission strategies, with reference to the previously sequenced SeMNPV strain from Florida (SeMNPV-US1) (IJkel *et al.*, 1999).

RESULTS AND DISCUSSION

Linking specific mutations to phenotypic variation is challenging because the phenotype of an individual derives from the specific assembly of the genetic information present at the whole-genome scale. Here, we attempted to infer the potential effects of the genomic variation observed in the genomes of European SeMNPV isolates. The phenotypes of these isolates, which differed in their transmission strategies, were characterized previously in terms of OB pathogenicity, speed of kill and OB yield, as well as their ability to induce sublethal infections in adults (Cabodevilla *et al.*, 2011). Our aim was to untangle the diversity of the mutations we observed in order to identify potential links between particular genome features and particular viral phenotypes.

Global genomic diversity in SeMNPV isolates

In this study, the genomes of eight SeMNPV isolates were compared. Six of them were from southern Spain (VT-SeAl1, VT-SeAl2, HT-SeG24, HT-SeG25, HT-SeG26 and HT-SeSP2A), one was from the UK (VT-SeOx4) and the reference isolate was from the USA (SeMNPV-US1). Annotation of the European genome sequences resulted in 137 putative ORFs (Table 1, Fig. 1), except for the HT-SeSP2A European reference isolate genome, which has a C→T mutation in position 23 399 producing a stop codon, splitting ORF Ac17 in two. There are several notable differences in ORF composition between the American SeMNPV-US1 (139 ORFs) (IJkel *et al.*, 1999) and European strain analysed in this report (137–138 ORFs). This difference results in part from the concatenation of two ORFs in European genomes compared with the SeMNPV-US1 genome. ORFs SeMNPV-US1-017 and SeMNPV-US1-018 resulted in only one ORF in the European genomes, which is homologous to lef-7 (Ac125). Similarly, ORFs SeMNPV-US1-022, SeMNPV-US1-023 and SeMNPV-US1-024 formed a single ORF, which is homologous to Hear105 in the European genomes. Two additional ORFs were also present in the European isolates. They are homologous to a RING finger cg30 protein and

Table 1. Assembly features and genomic variation between SeMNPV isolates

SeMNPV isolate	Genome length (bp)	No. of reads*	Sequence coverage	GenBank accession no.	Main distinguishing features
SeMNPV-US1	135 611	NA	NA	NC_002169	139 ORFs lef-7 splits in two Se17/18 Hear105 splits in three Se22/23/24
VT-SeAl1	135 653	50 142	130	HG425343	137 ORFs
VT-SeAl2	134 972	20 647	52	HG425344	137 ORFs
VT-SeOx4	142 709	22 549	57	HG425345	137 ORFs *piggyBac* transposon (7.7 kb) Truncated Maco35
HT-SeG24	135 292	23 774	60	HG425346	137 ORFs
HT-SeG25	135 556	11 106	26	HG425347	137 ORFs
HT-SeG26	135 718	15 235	34	HG425348	137 ORFs
HT-SeSP2A	135 395	19 413	46	HG425349	138 ORFs Ac17 splits in two Truncated odv-e66

*Number of reads assembled into each genome contig.

Maco134, present in the closely related *Spodoptera litura* nucleopolyhedrovirus II and *Mamestra configurata* nucleopolyhedrovirus B (MacoNPV-B), respectively.

All genomes were of similar length (~135 kbp), except VT-SeOx4, which had the largest genome (Table 1). Most of the size variation could be attributed to the presence of homologous repeat (hr) regions, principally in hr1, in which a ~600 bp insertion/deletion (indel) was identified. A global genome alignment revealed that the aligned nucleotide sequences were 97.3 % identical among all SeMNPV isolates and 98.7 % identical among the European isolates.

The larger genome size of VT-SeOx4 is due to a 7701 bp insertion between the ORFs Ld124 and cathepsin. BLAST searches revealed that this insertion contained three ORFs including a homologue of *Helicoverpa zea* single nucleopolyhedrovirus orf42 (HzSNPV42) and two pseudogenes: a *piggyBac* transposase and an ORF similar to the hypothetical protein LOC100570593 of the aphid, *Acyrthosiphon pisum*. The GC content of the insert (34.6 %) was lower than that observed in the rest of the VT-SeOX4 genome (43.9 %). This value was very similar to the 35.9 % GC content of the *Spodoptera frugiperda* genome (d'Alençon *et al.*, 2010). Given the presence of the TTAA motif on both sides of the insertion, this suggests that the 7.7 kb insert may have resulted from a *piggyBac*-mediated transposition event (Fig. 2). The transposon most likely originated from the genome of the host *S. exigua*. The frameshifts in the two main ORFs suggest that this transposon is no longer active.

Overall GC content values of 43.9 % were observed for all viral genomes, except for VT-SeOx4 in which the GC content was 43.4 %, due to the compositional bias of the transposable element. This value is also similar to that of other group II alphabaculoviruses, such as MacoNPV-A and MacoNPV-B (Li *et al.*, 2002a, b) (Table 1).

Mutation detection among European isolates

To assess potential links between genetic diversity and functional diversity, we identified the mutations, single nucleotide polymorphisms (SNPs) and indels present within the European SeMNPV population, as represented by seven isolates. In total, 419 SNPs were identified: 111 in non-coding regions and 309 in coding regions. With ~26.5 % of all SNPs detected in non-coding regions, an unexpectedly high prevalence of mutation was evident in non-coding regions, which represent only 9.3 % of the genome. Within coding regions, 200 mutations were synonymous substitutions and 109 SNPs were non-synonymous, resulting in amino acid changes. These changes in amino acid sequences were not distributed evenly in the genome. Of the 137 ORFs, 64 had no variation in deduced amino acid sequences and 10 ORFs had variation in repeat sequences or in low sequence coverage regions, causing statistical and analytical problems. Consequently, these were regarded as non-significant in our analyses even though they could have a functional impact. Mutations were observed in the remaining 63 ORFs: 51 had non-synonymous nucleotide substitutions, eight ORFs had both indels and SNPs, and four ORFs had only indels (Table 2).

Among the 109 non-synonymous SNPs, 60 were associated with changes in amino acid polarity, which might have a higher probability of altering protein function. Similarly, two were associated with changes to stop codons, resulting in frameshifts in ORFs Ac17 and odv-e66 of the HT-SeSP2A isolate (Table 2, Fig. 1). A total of 14 indels were identified in 12 ORFs, one of which resulted in a frameshift

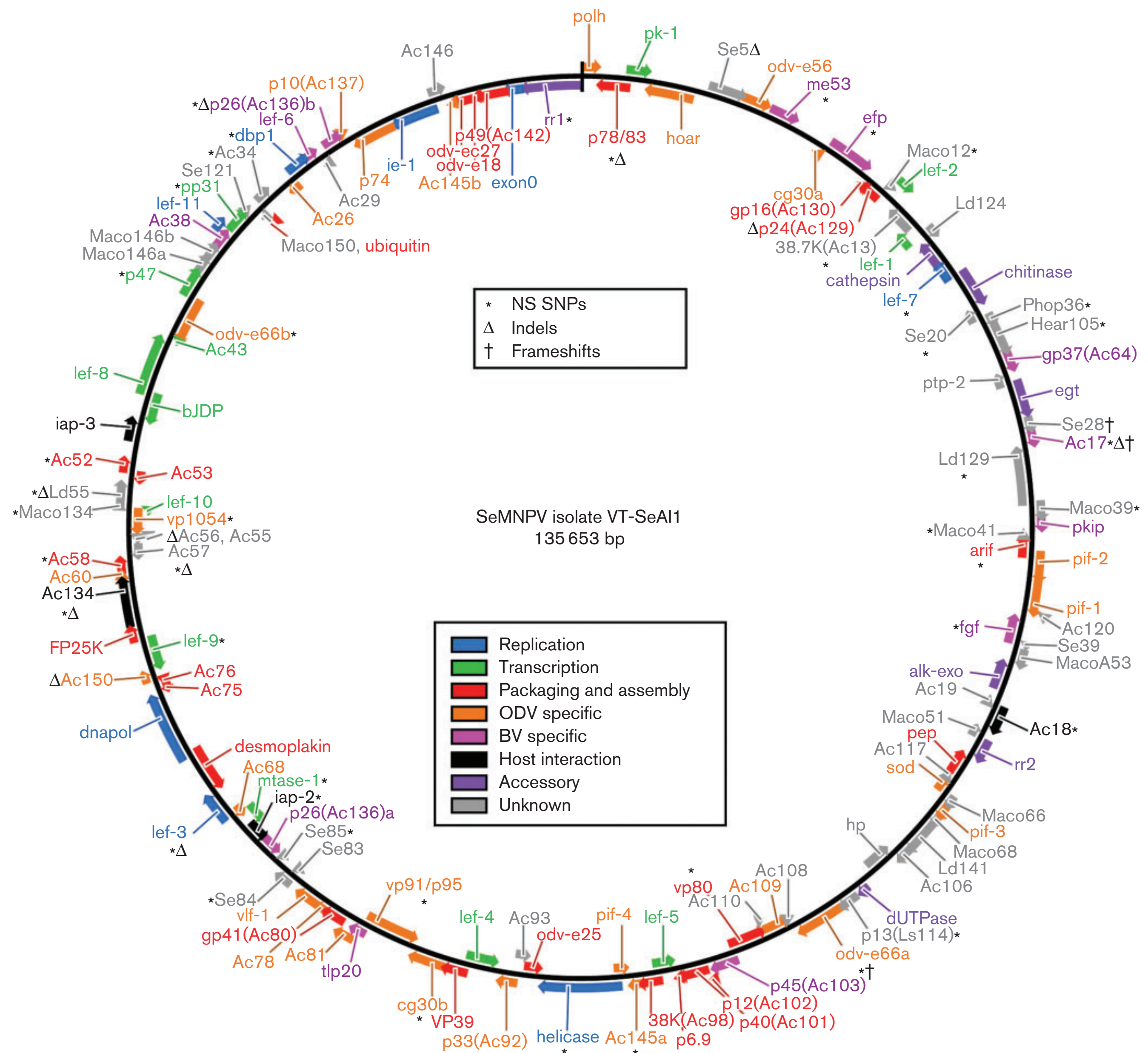

Fig. 1. Circular visualization of the annotation of the VT-SeAl1 isolate genome. The black circle represents the VT-SeAl1 genome; arrows above and below the genome represent sense and antisense ORFs, respectively. Each ORF is coloured according to the functional class. ORFs for which significant mutations have been found between European SeMNPV isolates are marked with symbols. NS, non-synonymous.

with the deletion of 4 bp in ORF Se28 of VT-SeOx4 and the truncation of 42 aa of the 190 aa protein. The other 13 indels were associated with single or several amino acid insertions or deletions in ORFs (Table 2, Fig. 1).

Overall, these 76 mutations were found in 48 ORFs belonging to all functional classes. Among these mutations, 31 were orphans found in only one genome, 23 were found in two out of the seven genomes and 22 were shared by three genomes (Table 2, Fig. 1).

Variation in non-coding regions

Non-coding regions of the SeMNPV genomes were observed to have accumulated a greater proportion of mutations than the coding regions, given the relative contributions of these regions to the total genome. Clearly, mutations in coding regions are likely to have deleterious effects, whereas in non-coding regions, such mutations are more likely to be neutral. However, certain non-coding regions could be important components of the genomes.

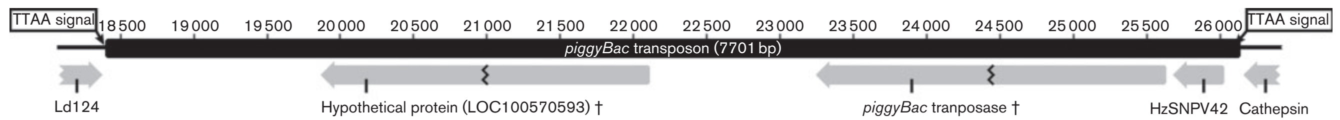

Fig. 2. *piggyBac* transposon from the genome of the VT-SeOx4 isolate. The 7701 bp insertion locates between Ld124 and cathepsin ORFs. The transposon contains two pseudogenes (†) and an intact ORF, and is flanked by TTAA insertion motifs.

For example, the homologous repeat regions are a well-recognized characteristic of nucleopolyhedrovirus genomes, but their sequence, number, length and position are specific to each virus (van Oers & Vlak, 2007). There is growing evidence that homologous repeat regions can function as potent enhancers of gene transcription (Lin *et al.*, 2010) or origins of DNA replication (Hilton & Winstanley, 2008; Hyink *et al.*, 2002; Vanarsdall *et al.*, 2007). The SeMNPV genome sequences showed variation in the size of homologous repeat regions, in particular hr1, among the European SeMNPV isolates. This had previously been suspected based on the examination of restriction endonuclease profile polymorphisms, which showed broad differences between genomes and allowed the localization of a 4–6 kb variable region in SeMNPV genomes from Spain (Cabodevilla *et al.*, 2011; Muñoz *et al.*, 1999). Variation in this region was principally due to the variable number of repeats of a 98 bp sequence in hr1, whereas this region was not present in the SeMNPV-US1 strain (Muñoz *et al.*, 1999). However, the functional importance of homologous repeat size variation resulting from indels remains elusive.

The SeOx4 genome contained a large insertion resulting from the transposition of a *piggyBac* element, which appeared to be non-functional (Fig. 1). As a result, this virus had the largest genome among our isolates. Unexpectedly, a significant correlation was observed between the size of the genome and the mean OB production per infected insect (OB per larva) (Spearmann coefficient=0.63) (Wessa, 2012). The largest genome VT-SeOx4 had the lowest productivity ($2.99\times10^8\pm4.59\times10^7$ OBs per larva), whereas the smallest genome VT-SeAl2 had the highest OB productivity ($1.14\times10^9\pm1.26\times10^8$ OBs per larva) (Cabodevilla *et al.*, 2011). This suggests that viral replication rate and production are somehow genome size dependent. This observation is in line with the appearance of defective genomes in cell culture systems (Giri *et al.*, 2012) and natural alphabaculovirus populations (Bull *et al.*, 2001; Simón *et al.*, 2005; Zwart *et al.*, 2008). However, as the European isolates of SeMNPV have variable genetic content, genome size is probably not the only factor modulating OB production among the isolates studied, as the virulence trade-off between speed of kill and yield has often been demonstrated in nucleopolyhedroviruses (Cory & Myers, 2003).

Notably, the insertion of transposable elements, like the *piggyBac* transposon present in the VT-SeOx4 isolate (Fig. 1), has been observed occasionally in other baculoviruses (Fraser *et al.*, 1995; Gilbert *et al.*, 2014; Jehle *et al.*, 1998). However, transposons do not seem to be maintained normally in baculovirus genomes, possibly due to the fitness costs imposed by the replication of larger genomes.

Variation in coding regions

In coding regions, an excess of synonymous mutations was detected, which do not result in amino acid changes, compared with non-synonymous mutations, which do result in amino acid changes and could potentially alter protein function. We assume that only frameshifts, indels and amino acid changes that are also associated with changes in amino acid polarity are most likely to be involved in functional changes, and emphasize those that are linked with marked phenotypic changes.

Variation in ORFs involved in DNA replication. Non-synonymous mutations were detected in six of the eight ORFs associated with baculovirus DNA replication (Table 2, Fig. 1). The mutations in the dnapol and ie-1 ORFs resulted in changes in specific amino acids, but not in their biochemical properties (polarity), and were therefore assumed to have little effect on protein function and virus phenotype. The dbp1, helicase, lef-3 and lef-7 ORFs harboured mutations associated with amino acid polarity changes. The SNP in dbp1 and the SNPs and the indel in lef-3 did not appear to be associated with any marked phenotype.

Remarkably, the SNPs in the helicase and lef-7 ORFs were only found in the HT-SeG25 isolate, for which the expression of the DNA polymerase in adult moths is highly reduced (Cabodevilla *et al.*, 2011). LEF-7 was shown to enhance baculovirus DNA replication and gene expression (Mitchell *et al.*, 2013). HT-SeG25 appears incapable of sustaining sublethal infection in adults and is poorly vertically transmitted, but efficiently transmitted horizontally. Mutations in the genes involved in DNA replication might affect the rate of virion production and thus might alter the spread of the infection within the host, thus affecting survival to adulthood. Another distinguishing feature of HT-SeG25 is that it is the most pathogenic isolate, requiring fewer OBs to establish a lethal infection (Cabodevilla *et al.*, 2011). If SNPs in the helicase and lef-7 ORFs result in more efficient DNA replication, more DNA could be packaged in the virions, which might lead to more occlusion-derived virions (ODVs) being incorporated in the OBs, resulting in OBs with greater insecticidal properties.

Table 2. Non-synonymous mutations observed in the SeMNPV European isolates

ORF name	Position	Mutation type	VT-SeAL1	Nucleotide change	Mutant strain	Change in amino acid polarity	Amino acid change	
p78/83	1370	SNP	T	C	VT-SEOx4	Yes	I	T
p78/83	1668	Indel	GACGACGTC	–	HT-SeG24, HT-SeSP2A	Yes	DVV	–
hoar	3703	SNP	A	G	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	No	L	P
Se5	6456	SNP	G	A	HT-SeG25, HT-SeG26, HT-SeSP2A	No	N	S
Se5	6482	Indel	ATG	–	HT-SeG25	Yes	M	–
Se5	6673	SNP	A	T	HT-SeG24, HT-SeSP2A	No	E	D
Se5	6770	Indel	GATGTCGACATCGATGCC	–	HT-SeG24	Yes	DVDIDA	–
Se5	6792	SNP	C	T	VT-SeAl2, VT-SEOx4, HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	No	A	V
Se5	7505	Indel	–	GAG	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	–	E
me53	9586	SNP	A	G	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	T	D
me53	9587	SNP	C	A	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	T	D
me53	9602	SNP	A	G	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	E	G
efp	12 914	SNP	T	C	HT-SeG26, HT-SeSP2A	Yes	S	F
efp	13 510	SNP	G	A	HT-SeSP2A	Yes	N	D
efp	14 278	SNP	A	G	HT-SeG25, HT-SeG26, HT-SeSP2A	No	I	V
efp	14 303	SNP	A	A	HT-SeG25, HT-SeG26, HT-SeSP2A	No	N	S
p24(Ac129)	15 169	Indel	–	GAC	HT-SeG25, HT-SeSP2A	Yes	–	V
Maco12	15 994	SNP	A	G	HT-SeG24, HT-SeG25	Yes	T	A
38.7K(Ac13)	17 088	SNP	A	T	HT-SeG26	No	S	T
38.7K(Ac13)	17 504	SNP	C	T	HT-SeG26	Yes	G	D
lef-7	20 664	SNP	C	T	HT-SeG25	Yes	D	N
Se20	23 224	SNP	G	A	HT-SeG24, HT-SeG25, HT-SeG26	Yes	L	S
Phop36	23 602	SNP	C	G	HT-SeG25, HT-SeG26	Yes	H	Q
Hear105	24 265	SNP	G	C	VT-SeAl2, VT-SEOx4, HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	P	R
Hear105	24 537	SNP	A	G	HT-SeG26	No	I	V
Hear105	24 667	SNP	G	A	HT-SeG26	Yes	D	G
Se28	29 099	Indel	CGAA	–	VT-SEOx4	Frameshift	RISKSSS	YQSRVAW
Ac17	29 634	SNP	C	T	HT-SeSP2A	Frameshift	R	*
Ac17	29 781	SNP	A	G	VT-SeAl2, VT-SEOx4, HT-SeG24, HT-SeG25, HT-SeSP2A	Yes	N	D
Ac17	29 823	Indel	–	CTCTCTCAGAAA	VT-SeAl2, VT-SEOx4, HT-SeG24, HT-SeG25, HT-SeSP2A	Yes	–	LSQK
Ld129	30 291	SNP	C	T	VT-SeAl2, VT-SEOx4, HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	No	R	H
Ld129	30 826	SNP	G	T	HT-SeG24, HT-SeG26	No	L	I
Ld129	31 752	SNP	G	T	VT-SeAl2, VT-SEOx4, HT-SeG24, HT-SeG25, HT-SeSP2A	Yes	T	K
Maco39	32 818	SNP	T	C	HT-SeG24, HT-SeSP2A	Yes	R	C
Maco39	33 155	SNP	A	G	HT-SeG24	No	N	S

Table 2. cont.

ORF name	Position	Mutation type	VT-SeAL1	Nucleotide change	Mutant strain	Change in amino acid polarity	Amino acid change	
Maco41	34 246	SNP	A	G	VT-SeAl2, VT-SEOx4, HT-SeG24	Yes	Y	H
arif	34 383	SNP	G	T	VT-SeAl2, VT-SEOx4, HT-SeG24	No	L	I
arif	34 785	SNP	G	A	VT-SeAl2, VT-SEOx4	Yes	C	R
fgf	38 617	SNP	C	A	VT-SeAl2, VT-SEOx4, HT-SeG25, HT-SeG26, HT-SeSP2A	No	D	E
fgf	38 774	SNP	G	A	VT-SeAl2, VT-SEOx4, HT-SeG25, HT-SeG26, HT-SeSP2A	No	V	A
fgf	38 922	SNP	C	T	VT-SeAl2, VT-SEOx4, HT-SeG25, HT-SeG26	Yes	D	N
alk-exo	41 071	SNP	A	T	HT-SeSP2A	No	L	M
Ac18	42 750	SNP	G	A	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	N	D
Ac18	43 102	SNP	C	T	HT-SeG24, HT-SeG25	Yes	T	M
Maco51	44 178	SNP	C	T	HT-SeG24	No	R	H
rr2	44 716	SNP	C	G	HT-SeG24	No	L	V
Maco66	47 762	SNP	C	G	HT-SeG26	No	L	V
Maco68	48 761	SNP	G	A	VT-SeAl2	No	N	S
Ld141	49 296	SNP	T	C	HT-SeG24, HT-SeG26	No	A	V
dUTPase	53 657	SNP	T	C	HT-SeG25, HT-SeG26, HT-SeSP2A	No	A	V
p13(Ls114)	54 066	SNP	C	G	VT-SeAl2, VT-SEOx4	Yes	R	G
odv-e66a	56 818	SNP	C	T	HT-SeSP2A	Frameshift	Q	*
odv-e66a	56 974	SNP	G	A	VT-SeAl2, VT-SEOx4, HT-SeG26	Yes	N	D
vp80	59 367	SNP	C	T	VT-SeAl2, VT-SEOx4, HT-SeSP2A	No	S	N
vp80	59 904	SNP	T	C	HT-SeG25, HT-SeSP2A	Yes	G	D
vp80	59 909	SNP	T	G	HT-SeG25, HT-SeSP2A	No	D	E
vp80	59 914	SNP	T	C	HT-SeG25, HT-SeSP2A	Yes	D	N
vp80	60 037	SNP	C	T	HT-SeG25, HT-SeSP2A	Yes	A	T
Ac145a	65 241	SNP	A	T	VT-SeAl2	No	S	C
Ac145a	65 355	SNP	A	G	VT-SeAl2, VT-SEOx4, HT-SeG24, HT-SeG26, HT-SeSP2A	Yes	N	D
helicase	67 004	SNP	C	G	HT-SeG25	Yes	N	K
helicase	68 312	SNP	C	A	HT-SeG24, HT-SeG25	No	E	D
lef-4	71 916	SNP	T	C	HT-SeG24	No	R	H
cg30b	75 079	SNP	T	C	HT-SeG25	Yes	P	S
cg30b	75 089	SNP	C	T	HT-SeG25	Yes	S	L
cg30b	75 239	SNP	A	C	VT-SeAl2, VT-SEOx4, HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	D	A
vp91/p95	76 424	SNP	A	G	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	No	L	P
vp91/p95	76 755	SNP	G	A	HT-SeG25, HT-SeG26	No	F	L
vp91/p95	77 219	SNP	T	C	HT-SeG25, HT-SeG26	Yes	R	Q
Se84	82 648	SNP	G	A	HT-SeG25	No	C	Y
Se84	83 082	SNP	G	T	HT-SeSP2A	Yes	A	S
Se85	83 554	SNP	T	C	HT-SeG25, HT-SeSP2A	No	S	N
Se85	83 642	SNP	T	G	HT-SeG24	Yes	H	N

Table 2. cont.

ORF name	Position	Mutation type	VT-SeAL1	Nucleotide change	Mutant strain	Change in amino acid polarity	Amino acid change	
iap-2	85 134	SNP	T	C	HT-SeG25	Yes	D	N
iap-2	85 239	SNP	A	G	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	W	R
iap-2	85 686	SNP	T	C	HT-SeG26	Yes	T	A
mtase-1	85 686	SNP	T	C	HT-SeG26	No	H	R
mtase-1	85 800	SNP	G	C	HT-SeG26	No	G	A
mtase-1	86 244	SNP	A	G	HT-SeG24, HT-SeSP2A	Yes	I	T
mtase-1	86 373	SNP	T	C	HT-SeG24, HT-SeSP2A	Yes	R	Q
mtase-1	86 425	SNP	C	T	HT-SeG24, HT-SeSP2A	Yes	E	K
lef-3	88 035	SNP	G	C	HT-SeSP2A	No	F	L
lef-3	88 039	Indel	GACGACGACATT	–	HT-SeSP2A	Yes	DDDI	–
lef-3	88 052	SNP	A	C	HT-SeSP2A	No	Y	S
lef-3	88 053	SNP	C	A	HT-SeSP2A	No	Y	S
lef-3	88 055	SNP	A	C	HT-SeSP2A	Yes	K	T
Desmoplakin	88 368	SNP	G	T	HT-SeG25, HT-SeSP2A	No	F	L
dnapol	92 851	SNP	C	T	HT-SeG25, HT-SeSP2A	No	L	F
Ac150	94 455	Indel	–	ATC	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	–	I
lef-9	94 738	SNP	T	C	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	D	N
Ac134	97 138	Indel	–	GTCGGCAGC	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	–	VGS
Ac134	98 704	SNP	G	A	VT-SeAl2, VT-SEOx4, HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	S	G
Ac58	99 744	SNP	G	A	HT-SeG24	Yes	T	A
Ac58	100 000	SNP	T	G	HT-SeG24, HT-SeSP2A	Yes	R	L
Ac58	100 144	SNP	C	G	HT-SeG24, HT-SeSP2A	No	A	G
Ac57	100 482	Indel	GTC	–	HT-SeG24, HT-SeG26, HT-SeSP2A	Yes	D	–
Ac57	100 488	SNP	C	A	HT-SeG24, HT-SeG26, HT-SeSP2A	Yes	C	G
Ac56	100 971	Indel	–	CGCTTC	VT-SeAl2, VT-SEOx4, HT-SeG26, HT-SeSP2A	Yes	–	EA
vp1054	102 328	SNP	T	A	HT-SeG24, HT-SeSP2A	Yes	F	Y
Maco134	102 636	SNP	T	A	HT-SeSP2A	Yes	H	L
Ld55	102 912	Indel	GCC	–	HT-SeG24, HT-SeG26, HT-SeSP2A	Yes	A	–
Ld55	102 925	SNP	T	A	HT-SeG24, HT-SeG26, HT-SeSP2A	Yes	D	V
Ld55	102 929	SNP	G	C	HT-SeG24, HT-SeG26, HT-SeSP2A	No	D	E
Ac52	104 429	SNP	C	T	VT-SEOx4	Yes	R	C
bJDB	107 867	SNP	A	G	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	No	V	A
odv-e66b	111 239	SNP	A	T	HT-SeG26	No	D	E
odv-e66b	111 706	SNP	G	C	HT-SeSP2A	Yes	R	G
odv-e66b	111 707	SNP	G	C	HT-SeSP2A	Yes	S	R
odv-e66b	111 711	SNP	A	T	HT-SeG25, HT-SeG26, HT-SeSP2A	Yes	L	H
odv-e66b	112 609	SNP	C	T	HT-SeG24, HT-SeG25, HT-SeSP2A	No	V	I
odv-e66b	113 107	SNP	T	G	HT-SeG26	Yes	H	N

Table 2. cont.

ORF name	Position	Mutation type	VT-SeAL1	Nucleotide change	Mutant strain	Change in amino acid polarity	Amino acid change	
p47	113 283	SNP	C	A	VT-SeAl2, VT-SEOx4	Yes	D	A
Maco146	115 186	SNP	T	A	HT-SeG25, HT-SeSP2A	No	L	F
pp31	117 184	SNP	A	G	HT-SeG24, HT-SeG26	No	I	V
pp31	117 569	SNP	G	A	VT-SeAl2, VT-SEOx4, HT-SeG26, HT-SeSP2A	Yes	D	G
Ac34	119 300	SNP	C	T	HT-SeG26	Yes	S	L
dbp1	121 221	SNP	G	C	VT-SeAl2, VT-SEOx4, HT-SeG25	Yes	Q	E
p26(Ac136)b	123 434	SNP	G	A	HT-SeG26	Yes	S	G
p26(Ac136)b	123 716	Indel	GACAAC	–	HT-SeG25, HT-SeG26	Yes	DN	–
ie-1	128 222	SNP	G	T	HT-SeG24, HT-SeG26, HT-SeSP2A	No	T	N
Ac145b	129 141	SNP	C	G	HT-SeG24, HT-SeSP2A	No	E	D
rr1	133 337	SNP	A	C	VT-SeAl2	Yes	L	R
rr1	134 229	SNP	A	T	HT-SeG24, HT-SeG25, HT-SeG26, HT-SeSP2A	No	S	T

*Refer to stop codons.

Variation in ORFs involved in viral transcription. Non-synonymous mutations were detected in six of the 13 ORFs associated with baculovirus transcription (Table 2, Fig. 1). The mutations in the lef-4 and dBJP ORFs resulted in changes in specifc amino acids, but not in their biochemical properties (polarity), and were therefore assumed to have little effect on protein function or virus phenotype. The mtase-1, lef-9, p47 and pp31 ORFs harboured mutations associated with amino acid polarity changes. The SNP in pp31 did not appear to be associated with any change in phenotype. The mtase-1 had three SNPs in the HT-SeG24 and HT-SeSP2A genomes, but did not seem to be correlated with any particular phenotype.

Notably, the two mutations in the subunits of the viral RNA polymerase, lef-9 and p47 ORFs are only present in the viruses associated with vertical transmission (VT-SeAl1, VT-SeAl2 and VT-SeOx4), which are also the least pathogenic. If these SNPs were to alter even slightly the rate of late gene transcription, they could result in the production of fewer ODVs occluded within OBs, which would then be less pathogenic.

Variation in ORFs involved with virion structure, packaging and assembly. The process of virion packaging, including nucleocapsid assembly, requires the interaction of numerous baculovirus proteins (Hou *et al.*, 2013). Furthermore, these viruses produce two types of virions: budded virions (BVs), which spread secondary infection within the host, and ODVs, which initiate the primary infection following consumption of OB-contaminated foliage by susceptible insects (horizontal transmission). Although both types of virions share a number of protein components, they also each comprise phenotype-specific proteins (Hou *et al.*, 2013). As changes in BVs or ODVs might lead to different virulence phenotypes, genes associated with virions were classified into one of three categories: (1) involved in the formation of both BVs and ODVs, (2) specific to ODVs and polyhedra, and (3) specific to BVs. Out of the 63 ORFs that contribute to the formation of baculovirus virions, non-synonymous mutations with the potential to alter protein function were detected in 21 ORFs (Table 2, Fig. 1).

ORFs involved in the formation of both BVs and ODVs. We found non-synonymous mutations in seven of the 24 ORFs involved in the formation of both BVs and ODVs (Table 2, Fig. 1). The mutations in desmoplakin changed the amino acid, but not its biochemical properties (polarity), and were therefore assumed to have little effect on protein function or virus phenotype. The p78/83, p24(Ac129), arif, vp80, Ac58 and Ac52 ORFs harboured mutations associated with amino acid polarity changes. The SNPs in arif and Ac58 did not appear to be associated with any clear phenotype. The vp80 ORF had three SNPs associated with the HT-SeG25 and HT-SeSP2A viruses, but these did not appear to alter the phenotype. Similarly, the indels in p78/83 and p24(Ac129) did not appear to be associated with any particular phenotype.

In contrast, the SNPs present in p78/83 and Ac52 were detected in VT-SeOX4, which is the fastest killing isolate with the lowest OB production per host. The p78/83 ORF is essential in lepidopteran baculoviruses and encodes a structural protein involved in nuclear actin assembly during baculovirus infection (Goley *et al.*, 2006). Ac52 has no determined function, but its disruption has multiple effects, including a dramatic reduction in BV production, and disruption of normal nucleocapsid envelopment and polyhedron formation, which results in decreased OB pathogenicity and an increase in time to death of infected insects (Tian *et al.*, 2009). The alteration of these ORFs may improve both virus primary infectivity and systemic infection, conferring a faster speed of kill, but a lower OB productivity.

ORFs specific to ODVs and polyhedra. Non-synonymous mutations were detected in nine of the 27 ORFs involved in the formation of ODVs and OBs (Table 2, Fig. 1). The mutations in Ac145b did not change amino acid polarity. ORFs odv-e66 (both copies), Ac145a, cg30b, vp91/95, Ac150 and vp1054 harboured mutations associated with amino acid polarity changes. The SNPs in odv-e66 (both copies), Ac145a and vp1054 do not appear to be associated with any particular phenotype. In *Autographa californica* multiple nucleopolyhedrovirus, cg30 was shown to confer a slight growth advantage to the virus (Passarelli & Miller, 1994). Two SNPs in cg30b are associated with the HT-SeG25 isolate, which is unable to establish persistent infection; however, their links with the phenotype are tenuous.

Interestingly three ORFs, hoar, vp91/95 and Ac150, had SNPs and indels bringing significant biochemical changes associated with the three viruses isolated following vertical transmission (VT-SeAl1, VT-SeAl2 and VT-SeOx4), which were also the less pathogenic isolates. The mutations in hoar and vp91/95 did not modify amino acid polarity, but as they involved a Leu to Pro change, they could bring an important conformation change to the proteins. There are also two SNPs in vp91/95 only present in HT-SeG25 and HT-SeG26, the most pathogenic isolates. Both vp91/95 and Ac150 have chitin-binding domains and are involved in primary infection of baculoviruses (Lapointe *et al.*, 2004; Rohrmann, 2011). Disruption of Ac150 results in a significant decrease in OB pathogenicity (Zhang *et al.*, 2005). The alteration of these ORFs may therefore affect the efficiency of primary infection through the midgut. Selection on this trait might be relaxed in vertically transmitted isolates.

ORFs specific to BVs. Non-synonymous mutations were present in five of the 12 ORFs involved in the formation of BVs (Table 2, Fig. 1). The me53, efp, Ac17, fgf and p26(Ac136)b (one of two copies) ORFs harboured mutations associated with amino acid polarity changes. The SNPs in efp, the SNPs and the indel in Ac17, and the SNPs in fgf and in p26(Ac136)b do not appear to be associated with any particular phenotype. The indel in p26(Ac136)b is only present in HT-SeG25 and HT-SeG26, the most pathogenic isolates. However, functional analyses (Simón *et al.*, 2008) revealed no differences when this indel was introduced and a correlation with pathogenic phenotype could not be inferred.

Remarkably, the three SNPs in me53 were only present in the three viruses isolated following vertical transmission (VT-SeAl1, VT-SeAl2 and VT-SeOx4), the least pathogenic isolates. This ORF encodes a protein containing a zinc-finger domain and is required for efficient BV production (de Jong *et al.*, 2009). The alteration of this ORF might improve BV production, conferring a better systemic infection in insect host cells and increasing baculovirus pathogenicity.

Variation in ORFs with accessory functions or involved with host interaction. Non-synonymous mutations were identified in seven of the 11 ORFs with accessory functions (Table 2, Fig. 1). The rr1, iap2, Ac18 and Ac134 ORFs harboured mutations associated with amino acid polarity changes. The SNP in rr1 is present only in VT-SeAl2, the least pathogenic and most productive isolate. The rr1 ORF is involved in DNA replication and repair by transforming RNA building blocks to DNA building blocks (Nordlund & Reichard, 2006). The alteration of this ORF could modulate OB productivity.

The SNPs in iap2 and Ac18 and the indel in Ac134 are present only in the three viruses isolated after vertical transmission (VT-SeAl1, VT-SeAl2 and VT-SeOx4), the least pathogenic isolates. Although its function is not well established, Ac18 may play a role in efficient virus infection that could increase the speed of kill (Wang *et al.*, 2007). Both iap2 and Ac134 are involved in the anti-apoptotic response (Clem & Miller, 1994; Friesen & Miller, 1987; Katsuma *et al.*, 2008). Slight changes in the inhibition of apoptosis might modulate the progression of systemic infection. This in turn could alter OB pathogenicity. Modulation of the anti-apoptotic response might also favour the establishment of covert infections in adults, favouring efficient vertical transmission.

Variation in ORFs with unknown functions. Finally, non-synonymous mutations were identified in 23 of the 42 ORFs of unknown function (Table 2, Fig. 1). Se5, Maco12, Ac13, Se20, Phop36, Hear105, Se28, Ld129, Maco39, Maco41, p13 (Ls114), Se84, Se85, Ac57, Ac56, Maco134, Ld55 and Ac34 harboured mutations associated with amino acid polarity changes. The SNPs and indels in most of these ORFs did not appear to be associated with any specific phenotype. However, the SNPs and indels in Se5, Se84 and Phop36 were associated with particular isolates, but functional analyses did not provide support for any particular phenotypic associations. However, the deletion of 4 bp in Se28 truncated approximately one-fifth of that protein in the VT-SeOx4 isolate. This isolate had a clear phenotype, as it was isolated from insects with established long-term covert infections in a laboratory

insect colony and was the fastest-killing virus isolate. Its mutated Se28 protein was likely non-functional, suggesting that the alteration of this ORF may have an important role in modulating sublethal infection in *S. exigua*.

CONCLUSIONS

This study of genomic diversity among several isolates of the alphabaculovirus SeMNPV revealed a high number of polymorphic sites. However, relatively few sites in the genomes were potentially involved in functional changes. Interestingly, these mutations were located in ORFs belonging to all functional classes. Changes in 13 ORFs (Ac18, Ac52, Ac134, Ac150, helicase, iap-2, lef-7, lef-9, me53, p47, p78/83, Se28 and vp91/95) could be correlated with particular phenotypic characteristics of the isolates. Notably, most of the changes identified could be involved in differences in transmission strategy. Mutations in two ORFs, the helicase and lef-7, are present in the HTSeG25 isolate, which appears unable to sustain covert infection required for vertical transmission. Mutations in three ORFs, Ac52, p78/83 and Se28, are found only in the VT-SeOX4 isolate, which has maintained long-term covert infection in a laboratory colony of *S. exigua*. Lastly, mutations in seven ORFs (Ac18, Ac134, Ac150, iap-2, lef-9, me53 and p47) are present only in the three isolates obtained following vertical transmission (VT-SeAl1, VT-SEAl2 and VT-SeOx4). Altogether, these ORFs, distributed among DNA replication, transcription, virus structure or host interactions, such as anti-apoptosis, suggest that baculovirus virulence or transmission phenotypes rely on the modulation of multiple molecular pathways. This study, unique to date for baculoviruses, provides a catalogue of potential lines of research based on clear predictions identified by us. Evidently, functional studies would need to be undertaken now to validate the phenotypic effect of the genomic variations that we identified as likely to be influential in specific virus phenotypic traits.

METHODS

Virus isolates. Seven isolates from our SeMNPV virus collection were used. Isolates HT-SeG24, HT-SeG25, and HT-SeG26 had been isolated from greenhouse soil substrates in Almería, Spain (Murillo *et al.*, 2007). The VT-SeAl1 and VT-SeAl2 isolates were obtained from virus-killed progeny of field-caught adults from Almerian greenhouses, whereas VT-SeOx4 originated from virus-killed progeny from a chronically infected laboratory population from Oxford, UK (Cabodevilla *et al.*, 2011). HT-SeSP2A had been subjected to detailed phenotypic characterization in previous studies and was included as the European reference isolate (Muñoz *et al.*, 1999). These isolates each comprised a single dominant genotype, as indicated by the absence of submolar fragments in restriction endonuclease profiles (Cabodevilla *et al.*, 2011).

Virus amplification and DNA isolation. In order to amplify each isolate, recently moulted *S. exigua* fourth instars were infected by the droplet feeding method (Hughes & Wood, 1981), using an OB concentration estimated previously to result in ~90 % mortality. Larvae were allowed to drink for a 10 min period or until they moved away from the droplet and were then transferred individually to a 24-compartment plate containing artificial diet. Larvae were reared at 25 ± 2 °C and cadavers were collected 5–7 days post-infection. OBs from virus-killed larvae were homogenized in 1 ml sterile water and filtered. DNA purification from OBs was performed by CsCl gradient to obtain DNA samples with a concentration $\geqslant 300$ ng μl^{-1} and $A_{260/280} \geqslant 1.8$.

DNA sequencing. Between 5 and 10 μg genomic DNA from each isolate was subjected to 454 sequencing (454 Lifesequencing). For each isolate, genome gap closing was obtained using the Sanger technique after PCR amplification of the non-overlapping regions using specific primers (Sistemas Genómicos).

Genome assembly and annotation. First, the reads of all isolates obtained by 454 sequencing together with the sequences from the Sanger technique were combined and then mapped onto the SeMNPV-US1 strain genome (IJkel *et al.*, 1999) using the runMapping program of the Newbler package (Margulies *et al.*, 2005). This mapping produced the 'European' genome consensus sequence (SeMNPV-EU), which was used subsequently as the assembly reference genome. Then, independent mappings of each isolate were performed using the SeMNPV-EU consensus, using the runMapping program, to obtain seven separate genome consensus sequences corresponding to VT-SeAl1, VT-SeAl2, VT-SeOx4, HT-SeG24, HT-SeG25, HT-SeG26 and HT-SeSP2A isolates. Following each mapping, consensus sequences were established, based on a combination of majority rule and read quality, and possible local assembly errors were manually corrected based on individual 454 reads.

Finally, ORF predictions were performed using the Glimmer3 program (Salzberg *et al.*, 1998) and ORF functional homologies were identified by aligning protein sequences of each ORF to the viral orthologous clusters of the Viral Bioinformatics Resource Center (Ehlers *et al.*, 2002; Upton *et al.*, 2003) and to the National Center for Biotechnology Information non-redundant protein database, using the BLASTP program (Altschul *et al.*, 1997). Each gene was assigned to a functional class, such as replication, transcription, virion assembly and packaging, BV specific, ODV specific, accessory or unknown based on literature (Hou *et al.*, 2013; Rohrmann, 2011).

Mutation detection. SNPs were detected using the SWP454 program (Brockman *et al.*, 2008) by mapping the reads of each isolate onto the VT-SeAl1 European isolate sequence, with a minimum of one read coverage. In case of polymorphic sequences within a single isolate, a 66 % consensus rule was applied by SWP454 regardless of read quality. Positions with indels were ignored in the SNP analysis. Indels were detected by comparing the results of codon-based multiple alignments of each ORF and including all isolates, using the CLUSTAL Omega program (Sievers *et al.*, 2011).

ACKNOWLEDGEMENTS

This study received financial support from the Spanish Ministry for Science and Technology (AGL2005-07909-CO3-01, AGL2008-05456-C03-01). O. C. received a Predoctoral Fellowship from the Spanish Ministry of Education and Culture. J. T. and E. A. H. are supported by the European Research Council (grant 205206 GENOVIR).

REFERENCES

Allen, D. E. & Little, T. J. (2009). Exploring the molecular landscape of host–parasite coevolution. *Cold Spring Harb Symp Quant Biol* **74**, 169–176.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z. Q., 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.

Bangham, J., Kim, K.-W., Webster, C. L. & Jiggins, F. M. (2008). Genetic variation affecting host–parasite interactions: different genes affect different aspects of sigma virus replication and transmission in *Drosophila melanogaster*. *Genetics* **178**, 2191–2199.

Belda, J. E., Mirasol, E., Escribano, A., Rapallo, S. & Caballero, P. (2000). [Efficacy of a nucleopolyhedrovirus (SeNPV) in the control of *Spodoptera exigua* (Hübner, 1808) (Lepidoptera: Noctuidae) in greenhouse pepper]. *Bol Sanid Veg, Plagas* **26**, 619–628 (in Spanish).

Bianchi, F. J. J. A., Joosten, N. N., Vlak, J. M. & van der Werf, W. (2001). The influence of greenhouse chrysanthemum on the interaction between the beet armyworm, *Spodoptera exigua*, and the baculovirus SeMNPV: parameter quantification for a process-based simulation model. *J Appl Entomol* **125**, 557–562.

Brockman, W., Alvarez, P., Young, S., Garber, M., Giannoukos, G., Lee, W. L., Russ, C., Lander, E. S., Nusbaum, C. & Jaffe, D. B. (2008). Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res* **18**, 763–770.

Bull, J. C., Godfray, H. C. & O'Reilly, D. R. (2001). Persistence of an occlusion-negative recombinant nucleopolyhedrovirus in *Trichoplusia ni* indicates high multiplicity of cellular infection. *Appl Environ Microbiol* **67**, 5204–5209.

Cabodevilla, O., Ibañez, I., Simón, O., Murillo, R., Caballero, P. & Williams, T. (2011). Occlusion body pathogenicity, virulence and productivity traits vary with transmission strategy in a nucleopolyhedrovirus. *Biol Control* **56**, 184–192.

Clem, R. J. & Miller, L. K. (1994). Control of programmed cell death by the baculovirus genes *p35* and *iap*. *Mol Cell Biol* **14**, 5212–5222.

Cory, J. S. & Myers, J. H. (2003). The ecology and evolution of insect baculoviruses. *Annu Rev Ecol Evol Syst* **34**, 239–272.

Costa, A., Jan, E., Sarnow, P. & Schneider, D. (2009). The Imd pathway is involved in antiviral immune responses in *Drosophila*. *PLoS ONE* **4**, e7436.

d'Alençon, E., Sezutsu, H., Legeai, F., Permal, E., Bernard-Samain, S., Gimenez, S., Gagneur, C., Cousserans, F., Shimomura, M. & other authors (2010). Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. *Proc Natl Acad Sci U S A* **107**, 7680–7685.

de Jong, J., Arif, B. M., Theilmann, D. A. & Krell, P. J. (2009). *Autographa californica* multiple nucleopolyhedrovirus *me53* (*ac140*) is a nonessential gene required for efficient budded-virus production. *J Virol* **83**, 7440–7448.

Ehlers, A., Osborne, J., Slack, S., Roper, R. L. & Upton, C. (2002). Poxvirus orthologous clusters (POCs). *Bioinformatics* **18**, 1544–1545.

Erlandson, M. A. (2009). Genetic variation in field populations of baculoviruses: mechanisms for generating variation and its potential role in baculovirus epizootiology. *Virol Sin* **24**, 458–469.

Fraser, M. J., Cary, L., Boonvisudhi, K. & Wang, H. G. (1995). Assay for movement of Lepidopteran transposon IFP2 in insect cells using a baculovirus genome as a target DNA. *Virology* **211**, 397–407.

Friesen, P. D. & Miller, L. K. (1987). Divergent transcription of early 35- and 94-kilodalton protein genes encoded by the *Hin*dIII K genome fragment of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J Virol* **61**, 2264–2272.

Gilbert, C., Chateigner, A., Ernenwein, L., Barbe, V., Bézier, A., Herniou, E. A. & Cordaux, R. (2014). Population genomics supports baculoviruses as vectors of horizontal transfer of insect transposons. *Nat Commun* **5**, 3348.

Giri, L., Feiss, M. G., Bonning, B. C. & Murhammer, D. W. (2012). Production of baculovirus defective interfering particles during serial passage is delayed by removing transposon target sites in *fp25k*. *J Gen Virol* **93**, 389–399.

Goley, E. D., Ohkawa, T., Mancuso, J., Woodruff, J. B., D'Alessio, J. A., Cande, W. Z., Volkman, L. E. & Welch, M. D. (2006). Dynamic nuclear actin assembly by Arp2/3 complex and a baculovirus WASP-like protein. *Science* **314**, 464–467.

Gröner, A. (1986). Specificity and safety of baculoviruses. In *The Biology of Baculoviruses: Biological Properties and Molecular Biology*, pp. 177–202. Edited by R. R. Granados & B. A. Federici. San Diego, CA: Academic Press.

Herniou, E. A., Arif, B. M., Becnel, J. J., Bissard, G. W., Bonning, B., Harrison, R., Jehle, J. A., Theilmann, D. A. & Vlak, J. M. (2011). Baculoviridae. In *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, pp. 163–173. Edited by A. M. King, E. Lefkowitz, M. J. Adams & E. B. Carstens. London: Academic Press.

Hilton, S. & Winstanley, D. (2008). Genomic sequence and biological characterization of a nucleopolyhedrovirus isolated from the summer fruit tortrix, *Adoxophyes orana*. *J Gen Virol* **89**, 2898–2908.

Hou, D., Zhang, L., Deng, F., Fang, W., Wang, R., Liu, X., Guo, L., Rayner, S., Chen, X. & other authors (2013). Comparative proteomics reveal fundamental structural and functional differences between the two progeny phenotypes of a baculovirus. *J Virol* **87**, 829–839.

Hughes, P. R. & Wood, H. A. (1981). A synchronous peroral technique for the bioassay of insect viruses. *J Invertebr Pathol* **37**, 154–159.

Hyink, O., Dellow, R. A., Olsen, M. J., Caradoc-Davies, K. M. B., Drake, K., Herniou, E. A., Cory, J. S., O'Reilly, D. R. & Ward, V. K. (2002). Whole genome analysis of the *Epiphyas postvittana* nucleopolyhedrovirus. *J Gen Virol* **83**, 957–971.

IJkel, W. F. J., van Strien, E. A., Heldens, J. G. M., Broer, R., Zuidema, D., Goldbach, R. W. & Vlak, J. M. (1999). Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *J Gen Virol* **80**, 3289–3304.

Jehle, J. A., Nickel, A., Vlak, J. M. & Backhaus, H. (1998). Horizontal escape of the novel Tc1-like lepidopteran transposon TCp3.2 into *Cydia pomonella* granulovirus. *J Mol Evol* **46**, 215–224.

Katsuma, S., Kawaoka, S., Mita, K. & Shimada, T. (2008). Genome-wide survey for baculoviral host homologs using the *Bombyx* genome sequence. *Insect Biochem Mol Biol* **38**, 1080–1086.

Kolodny-Hirsch, D. M., Sitchawat, T., Jansiri, T., Chenrchaivachirakul, A. & Ketunuti, U. (1997). Field evaluation of a commercial formulation of the *Spodoptera exigua* (Lepidoptera: Noctuidae) nuclear polyhedrosis virus for control of beet armyworm on vegetable crops in Thailand. *Biocontrol Sci Technol* **7**, 475–488. .

Lapointe, R., Popham, H. J. R., Straschil, U., Goulding, D., O'Reilly, D. R. & Olszewski, J. A. (2004). Characterization of two *Autographa californica* nucleopolyhedrovirus proteins, Ac145 and Ac150, which affect oral infectivity in a host-dependent manner. *J Virol* **78**, 6439–6448.

Lasa, R., Ruiz-Portero, C., Alcázar, M. D., Belda, J. E., Caballero, P. & Williams, T. (2007). Efficacy of optical brightener formulations of *Spodoptera exigua* multiple nucleopolyhedrovirus (SeMNPV) as a biological insecticide in greenhouses in southern Spain. *Biol Control* **40**, 89–96.

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

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

Lin, X., Chen, Y., Yi, Y. & Zhang, Z. Q. (2010). Baculovirus immediately early 1, a mediator for homologous regions enhancer function *in trans*. *Virol J* **7**, 32.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y.-J. & other authors (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

Martignoni, M. E. & Iwai, P. J. (1981). A catalogue of viral diseases of insects, mites and ticks. In *Microbial Control of Pests and Plant Diseases 1970–1980*, pp. 897–911. Edited by H. D. Burges. London: Academic Press.

Mitchell, J. K., Byers, N. M. & Friesen, P. D. (2013). Baculovirus F-box protein LEF-7 modifies the host DNA damage response to enhance virus multiplication. *J Virol* **87**, 12592–12599.

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

Muñoz, D., Castillejo, J. I. & Caballero, P. (1998). Naturally occurring deletion mutants are parasitic genotypes in a wild-type nucleopolyhedrovirus population of *Spodoptera exigua*. *Appl Environ Microbiol* **64**, 4372–4377.

Muñoz, D., Murillo, R., Krell, P. J., Vlak, J. M. & Caballero, P. (1999). Four genotypic variants of a *Spodoptera exigua* nucleopolyhedrovirus (Se-SP2) are distinguishable by a hypervariable genomic region. *Virus Res* **59**, 61–74.

Murillo, R., Muñoz, D., Ruíz-Portero, M. C., Alcázar, M. D., Belda, J. E., Williams, T. & Caballero, P. (2007). Abundance and genetic structure of nucleopolyhedrovirus populations in greenhouse substrate reservoirs. *Biol Control* **42**, 216–225.

Nordlund, P. & Reichard, P. (2006). Ribonucleotide reductases. *Annu Rev Biochem* **75**, 681–706.

Passarelli, A. L. & Miller, L. K. (1994). *In vivo* and *in vitro* analyses of recombinant baculoviruses lacking a functional *cg30* gene. *J Virol* **68**, 1186–1190.

Rohrmann, G. F. (2011). *Baculovirus Molecular Biology*, 2nd edn. Bethesda, MD: National Center for Biotechnology Information.

Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* **26**, 544–548.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M. & other authors (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539.

Simón, O., Williams, T., López-Ferber, M. & Caballero, P. (2004). Virus entry or the primary infection cycle are not the principal determinants of host specificity of *Spodoptera* spp. nucleopolyhedroviruses. *J Gen Virol* **85**, 2845–2855.

Simón, O., Williams, T., López-Ferber, M. & Caballero, P. (2005). Functional importance of deletion mutant genotypes in an insect nucleopolyhedrovirus population. *Appl Environ Microbiol* **71**, 4254–4262.

Simón, O., Williams, T., Caballero, P. & Possee, R. D. (2008). Effects of *Acp26* on *in vitro* and *in vivo* productivity, pathogenesis and virulence of *Autographa californica* multiple nucleopolyhedrovirus. *Virus Res* **136**, 202–205.

Smits, P. H., van de Vrie, M. & Vlak, J. M. (1987). Nuclear polyhedrosis virus for control of *Spodoptera exigua* larvae on glasshouse crops. *Entomol Exp Appl* **43**, 73–80.

Tian, C.-H., Zhao, J.-F., Xu, Y.-P., Xue, J., Zhang, B.-Q., Cui, Y.-J., Zhang, M.-J., Bao, Y.-Y. & Zhang, C.-X. (2009). Involvement of *Bombyx mori* nucleopolyhedrovirus ORF41 (Bm41) in BV production and ODV envelopment. *Virology* **387**, 184–192.

Upton, C., Slack, S., Hunter, A. L., Ehlers, A. & Roper, R. L. (2003). Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. *J Virol* **77**, 7590–7600.

van Oers, M. M. & Vlak, J. M. (2007). Baculovirus genomics. *Curr Drug Targets* **8**, 1051–1068.

Vanarsdall, A. L., Mikhailov, V. S. & Rohrmann, G. F. (2007). Baculovirus DNA replication and processing. *Curr Drug Targets* **8**, 1096–1102.

Wang, Y., Wu, W., Li, Z., Yuan, M., Feng, G., Yu, Q., Yang, K. & Pang, Y. (2007). *ac18* is not essential for the propagation of *Autographa californica* multiple nucleopolyhedrovirus. *Virology* **367**, 71–81.

Wessa, P. (2012). Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7. http://www.wessa.net/.

Wilfert, L. & Jiggins, F. M. (2010). Host–parasite coevolution: genetic variation in a virus population and the interaction with a host gene. *J Evol Biol* **23**, 1447–1455.

Zhang, J.-H., Ohkawa, T., Washburn, J. O. & Volkman, L. E. (2005). Effects of *Ac150* on virulence and pathogenesis of *Autographa californica* multiple nucleopolyhedrovirus in noctuid hosts. *J Gen Virol* **86**, 1619–1627.

Zwart, M. P., van Oers, M. M., Cory, J. S., van Lent, J. W. M., van der Werf, W. & Vlak, J. M. (2008). Development of a quantitative real-time PCR for determination of genotype frequencies for studies in baculovirus population biology. *J Virol Methods* **148**, 146–154.