Journal of Invertebrate Pathology 107 (2011) 33–42

Contents lists available at ScienceDirect

Journal of Invertebrate Pathology

journal homepage: www.elsevier.com/locate/jip

Sequence comparison between three geographically distinct *Spodoptera frugiperda* multiple nucleopolyhedrovirus isolates: Detecting positively selected genes ☆

Oihane Simón [a], Leopoldo Palma [a], Inés Beperet [a], Delia Muñoz [b], Miguel López-Ferber [c], Primitivo Caballero [a,b], Trevor Williams [d,*]

[a] Instituto de Agrobiotecnología, CSIC, Universidad Pública de Navarra, Gobierno de Navarra, Campus Arrosadia, 31192 Mutilva Baja, Navarra, Spain
[b] Laboratorio de Entomología Agrícola y Patología de Insectos, Universidad Pública de Navarra, 31006 Pamplona, Spain
[c] Ecole des Mines d'Alès, 6 avenue de Clavières, F. 30319 Alès Cedex, France
[d] Instituto de Ecología AC, Apartado Postal 63, Xalapa, Veracruz 91070, Mexico

ARTICLE INFO

Article history:
Received 28 July 2010
Accepted 10 January 2011
Available online 14 January 2011

Keywords:
Spodoptera frugiperda
Nucleopolyhedrovirus
Nicaraguan isolate
Genomic sequence
Positive selected genes

ABSTRACT

The complete genomic sequence of a Nicaraguan plaque purified *Spodoptera frugiperda* nucleopolyhedrovirus (SfMNPV) genotype SfMNPV-B was determined and compared to previously sequenced isolates from United States (SfMNPV-3AP2) and Brazil (SfMNPV-19). The genome of SfMNPV-B (132,954 bp) was 1623 bp and 389 bp larger than that of SfMNPV-3AP2 and SfMNPV-19, respectively. Genome size differences were mainly due to a deletion located in the SfMNPV-3AP2 *egt* region and small deletions and point mutations in SfMNPV-19. Nucleotide sequences were strongly conserved (99.35% identity) and a high degree of predicted amino acid sequence identity was observed. A total of 145 open reading frames (ORFs) were identified in SfMNPV-B, two of them (*sf39a* and *sf110a*) had not been previously identified in the SfMNPV-3AP2 and SfMNPV-19 genomes and one (*sf57a*) was absent in both these genomes. In addition, *sf6* was not previously identified in the SfMNPV-19 genome. In contrast, SfMNPV-B and SfMNPV-19 both lacked *sf129* that had been reported in SfMNPV-3AP2. In an effort to identify genes potentially involved in virulence or in determining population adaptations, selection pressure analysis was performed. Three ORFs were identified undergoing positive selection: *sf49* (*pif-3*), *sf57* (*odv-e66b*) and *sf122* (unknown function). Strong selection for ODV envelope protein genes indicates that the initial infection process in the insect midgut is one critical point at which adaptation acts during the transmission of these viruses in geographically distant populations. The function of ORF *sf122* is being examined.

1. Introduction

The fall armyworm, *Spodoptera frugiperda* (JE Smith) (Lepidoptera: Noctuidae) is a severe pest of maize, sorghum and rice in tropical and subtropical areas of the Americas. Isolates of *S. frugiperda* multiple nucleopolyhedrovirus (SfMNPV) are known to cause natural epizootics of disease and have been recognized as potential control agents against this pest (Escribano et al., 1999; Fuxa, 1982; Shapiro et al., 1991). A Nicaraguan isolate (SfMNPV-NIC) was selected for formulation and field trails in Honduras and Mexico (Williams et al., 1999), and has been extensively studied at the genomic and phenotypic levels (López-Ferber et al., 2003; Simón et al., 2004, 2005b, 2008c). SfMNPV-3AP2 is a single genotype isolate from Missouri (USA) with a fast killing phenotype that was tentatively attributed to its lack of the *egt* gene (Harrison et al., 2008). The SfMNPV-19 isolate, purified from a single larval cadaver collected in Paramá State, southern Brazil, was selected from 22 SfMNPV isolates collected in Brazil due to its high pathogenicity against fall armyworm larvae (Barreto et al., 2005; Wolff et al., 2008).

Adaptation of viruses to their local host populations has been described in various occasions. Comparative studies have reported that a Honduran population of *S. frugiperda* was most susceptible to a Central American isolate (Escribano et al., 1999), whereas insects from a Colombian population were most susceptible to a SfMNPV isolate from Colombia (O. Simón, unpublished data). Similar findings have been observed by others working on different insect-nucleopolyhedrovirus systems elsewhere (Erlandson et al., 2007; Erlandson, 2009). This pattern is likely to be generated through processes of host–pathogen co-evolution, in which selection favors viruses with high transmissibility in the local host population. A consequence of this is that the choice of an isolate as the basis for a biological insecticide requires the assessment of a range of geographical isolates, including those from the region in which the biological control program is required.

☆ The GenBank/EMBL/DDBJ accession number of the sequence reported in this paper is HM595733.

* Corresponding author. Fax: +52 228 818 78 09.
E-mail address: trevor.inecol@gmail.com (T. Williams).

0022-2011/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jip.2011.01.002

A number of genes intervene in the species-specific virulence or host range of NPVs by affecting the ability to infect and replicate in cells and thereby influencing the dose required for host mortality, or the survival time of infected larvae from a given host population or species (Chen and Thiem, 1997; Chen et al., 1998; Popham et al., 1998). In most cases, species-specific effects of a gene on virulence or host range have been elucidated by ORF disruption or deletion. Eliminating or reducing virulence against one species following knock-out of a particular gene suggests that acquisition of new genes via recombination has the potential to shape NPV host range. However, very small changes such as individual amino acid replacements in particular viral genes have also proven influential in expanding host range or modifying virulence (Argaud et al., 1998; Kamita and Maeda, 1997; Yang, 1998). Additionally, nucleotide substitutions may lead to alterations in the activity of the encoded protein that facilitate adaptation to a new host or effectively overcome the defenses of a current host. Such mutations would confer a selective advantage and would be fixed in the population at a higher rate than silent or neutral substitutions (Harrison and Bonning, 2004).

The availability of the complete sequences of isolates from the same virus species from different geographical areas offers the possibility of another kind of analysis, based on a bioinformatics approach. In the present study, the complete sequence of an SfMNPV-NIC genotype was compared with those of SfMNPV-3AP2 and SfMNPV-19. Maximum-likelihood models of codon substitutions were used to test for selection on ORFs. The non-synonymous to synonymous substitution rate ratio ($\omega = d_N/d_S$) provides a sensitive measure of selection on amino acid sequences. We reasoned that comparative analyses of SfMNPV genomes from geographically distinct sources could highlight genetic differences involved in the marked phenotypic variation in virulence that has been reported in these isolates, and thereby focus attention on the key genes involved in traits that are considered desirable for the development of effective bioinsecticidal products.

2. Materials and methods

2.1. Virus

The Nicaraguan SfMNPV isolate (SfMNPV-NIC) was initially collected from larval cadavers collected from a maize field in Nicaragua. The isolate was propagated in *S. frugiperda* larvae reared in the laboratory and subjected to an uncertain number of passages. SfMNPV-NIC genotypic variants were isolated by plaque purification (Simón et al., 2004) and the predominant genotype that possessed the largest genome (SfMNPV-B) was selected for bacmid construction (Simón et al., 2008a).

Occlusion bodies (OBs) obtained from larvae infected with SfMNPV-B were extracted filtered through cheesecloth, washed twice with 0.1% SDS and once with 0.1 M NaCl and finally resuspended in bidistilled water. OB suspensions were quantified using a bacterial counting chamber and stored at 4 °C until used.

2.2. Viral DNA isolation and cloning

SfMNPV-B bacmid DNA was sequenced to ensure that the nucleotide sequence originated from a single genotype and was not the result of contamination from a persistent infection present in the insect colony (Simón et al., 2010). The complete SfMMPV-B genome was cloned in a pBACe3.6 vector that includes a chloramphenicol resistance gene (Frengen et al., 1999). The pUC19 plasmid was removed from pBACe3.6 by treatment with *Bam*HI. pBluescript KS-I with a modified polylinker that included the *Asc*I restriction site, was inserted using T4 DNA ligase (New England Biolabs, Ipswich, MA). To include the *Asc*I restriction site in pBluescript KS-I polylinker, the polylinker was amplified by PCR using two specific primers. One of the primers included a *Sac*I and an *Asc*I restriction site sequence, and a 20 nt homolog to the right arm of the polylinker, whereas the second primer included a 20 nt homolog to the left arm of the polylinker, and *Asc*I and *Kpn*I restriction site sequences. The PCR product was double digested with *Sac*I and *Kpn*I and cloned into pBluescript KS-I previously digested with these two enzymes. Once the modified pBACe3.6 was obtained, this and the SfMNPV-B DNA purified by CsCl were digested with *Asc*I and ligated overnight at 16 °C using the T4 DNA ligase (Simón et al., 2008a). The ligation reaction was dialyzed for 4 h against TE buffer. After dialysis, the ligation was used to transform DH10B GeneHogs electrocompetent cells. Transformed cells were incubated at 37 °C for 1 h in SOC medium and colonies were selected in the presence of chloramphenicol. Colonies were selected on chloramphenicol medium and bacmid DNAs were purified by alkaline lysis. DNA was digested with *Pst*I for comparison with the restriction profiles of each of the Nicaraguan SfMNPV genotypes (Simón et al., 2004). The bacmid with the complete genotype was selected for sequencing. Alkaline lysis and CsCl gradient purified bacmid DNA was used for sequencing (Sistemas Genómicos S.L., Paterna, Valencia, Spain).

2.3. DNA sequencing and sequence analysis

Shotgun sequencing was performed using a genomic library of SfMNPV-B bacmid DNA constructed into a sequencing vector. Sequence information was generated from 2304 reactions performed on 1152 clones. The depth of sequence coverage across the genome was 4–8x. Sequencing reactions were set up using the ABI Prism Big Dye Terminator Cycle Sequencing Reaction kit on a 9600 of PE model thermocycler. The reaction products were loaded in an automated DNA sequencer ABI PRISM. The DNA and deduced amino acid sequences were compared against the updated GenBank/EMBL, SWISS-PROT and PIR databases using BLASTn, (Altschul et al., 1990; Pearson, 1990). SfMNPV-3AP2 (accession number EF035042) and SfMNPV-19 (accession number EU258200) genomic sequences were obtained from the GenBank database and published sources. Sequence alignments and gene-parity plots were performed using NCBI BLAST alignment tools (Altschul et al., 1990) to examine genome organization and order of homologous ORFs. SignalP (Bendtsen et al., 2004), TargetP (Emanuelsson et al., 2000) and Virus-Ploc tools (Chou and Shen, 2008) were also used to determine the nature of selected proteins of unknown function.

2.4. Analysis of selection pressures

PAML software package version 4 (Yang, 2007; http://abacus.gene.ucl.ac.uk/software/paml.html) was used to estimate positive selection acting on SfMNPV ORFs and to infer amino acid sites subject to this process. This software uses a maximum-likelihood approach to determine the numbers of non-synonymous (amino acid changing) substitutions per non-synonymous site (d_N) and of synonymous (silent) substitutions per synonymous site (d_S). The ratio of d_N to d_S, ω, is a measure of the magnitude of selection acting on a gene. Genes with a value of $\omega = 1$ are deemed to be subjected to neutral selection, so that that non-synonymous mutations have no effect on fitness. Genes with $\omega < 1$ are undergoing negative or purifying selection, in which non-synonymous mutations are eliminated at a faster rate than synonymous mutations because of their deleterious influence on fitness. Finally, genes for which $\omega > 1$ are undergoing positive or diversifying selection in which non-synonymous mutations are fixed at a faster rate than synonymous mutations as they positively influence fitness. To achieve this, sequences were examined for the presence

of indels causing a frameshift and the affected codons were removed from the alignments; indels that did not result in a frameshift were not eliminated. Sequence data were then fitted to codon-based substitution models that allow ω to vary among sites by using the CODEML application implemented in PAML. In the simplest model M0 (one ratio), it is assumed that the ω ratio is an average over all sites. The "nearly neutral" model (M1) allows for conserved sites when $0 < \omega < 1$ and completely neutral sites when $\omega = 1$. The "positive selection model" (M2) adds a third class to M1 in which ω can take values >1. Model M3 (discrete) has three classes with proportions p_0, p_1 and p_2 and ω_0, ω_1 and ω_2 values estimated from the data. Model M7 (β) assumes a beta continuous probability distribution and does not allow for sites with $\omega > 1$, model 8 (β and ω) adds an extra class of sites to the M7 model in which ω can be greater than one. Models M0 and M1 are nested with models M2 and M3, and model M7 is nested with model M8. Nested models can prove particularly useful since log-likelihood values can be compared using a likelihood ratio test. The level of significance was calculated as follows: the difference between the log-likelihood values for two models was doubled and the resulting value was compared with a χ^2 distribution with the number of degrees of freedom calculated from the difference in the number of parameters between the models. The comparison provides a *P* value that allows the null hypothesis (no positive selection, integrated in models M0, M1 and M7) to be accepted or rejected. Positive selection can be only inferred when models M2, M3 or M8 indicate codons with a ω ratio >1 and the likelihood ratio test of positive selection is significant at $P < 0.05$. Finally, a Bayesian Empirical Bayes (BEB) approach was used to identify the sites that potentially are under positive selection in the protein. Thus, identification of genes with these values represents persuasive evidence of adaptive evolution.

3. Results and discussion

3.1. General characteristics of the Nicaraguan SfMNPV genome

The complete genome of the SfMNPV-B plaque isolate comprised 132,954 bp, approximately 4 kb larger than the genome size estimated for this isolate by REN analysis (Simón et al., 2005a), and 1617 and 389 bp larger than the SfMNPV-3AP2 and SfMNPV-19 genomes, respectively. These size differences among geographically distinct SfMNPV isolates were accounted for almost entirely by a deletion of 1428 bp located in the *egt* region of the SfMNPV-3AP2 genome (Harrison et al., 2008) and by a series of point mutations and short deletions or insertions that resulted in the addition or deletion of codons in the SfMNPV-19 genome (Wolff et al., 2008). The SfMNPV-B genome sequence has a G + C content of 40.30% compared to 40.20% and 40.26% for SfMNPV-3AP2 and SfMNPV-19, respectively (Harrison et al., 2008; Wolff et al., 2008). This value is also comparable to those of other group II NPVs, including *Spodoptera exigua* MNPV (SeMNPV) (Ijkel et al., 1999) or *Mamestra configurata* MNPV B (MacoNPV-B) (Li et al., 2002).

The computationally-derived SfMNPV-B *Pst*I restriction map was in agreement with the experimentally constructed physical map from REN analysis performed previously for this isolate (Simón et al., 2005a). However differences were observed in the number and position of restriction sites along the SfMNPV-3AP2 and SfMNPV-19 genomes, that resulted in *Pst*I profiles that differed for each isolate. The overall nucleotide sequence identity among the different isolates (~99.35%) was extremely high and alignment of the genomic sequences from SfMNPV-B, SfMNPV-3AP2 and SfMNPV-19 viruses with Clustal W confirmed that the three isolates clearly represent strains of the same virus species. The 1428 bp deletion of the SfMNPV-3AP2 genome partially overlaps the *egt* and *sf27* ORFs. Their size is reduced to only 207 and 81 codons, respectively, whereas the *egt* and *sf27* genes of the SfMNPV-B genome are 525 and 168 codons respectively (Supplemental material, Table S1), in agreement with findings for the Brazilian isolate (Tumilasci et al., 2003; Wolff et al., 2008). The baculovirus *egt* encodes an ecdysteroid UDP-glucosyltransferase (EGT) that alters the hormonal balance of the host larva, preventing the larval molt and ensuring the continuation of insect feeding and weight gain (ÓReilly, 1995; ÓReilly and Miller, 1991). The continued growth of the infected insect is reflected in an increased production of viral OBs. Deletion of *egt* from the viral genome has been shown to increase the speed of kill of the virus with a corresponding reduction in crop damage (Chen et al., 2000; ÓReilly, 1995; Slavicek et al., 1999; Treacy et al., 1997). Harrison et al. (2008) suggested that the deletion of most of the *egt* ORF in SfMNPV-3AP2 provided a potential explanation for its rapid speed of kill in bioassays. However, minority genotypes present in the Nicaraguan SfMNPV isolate that present deletions that encompass *egt*, among other genes, proved to be less virulent than deleted genotypes that carried this gene intact, possibly due to the absence of genes located elsewhere (Simón et al., 2004, 2005b). This deletion also affected the *sf27* ORF of SfMNPV-3AP2 (Table S1), an ORF that occurs only in group II NPVs (Harrison et al., 2008; Ijkel et al., 1999; Jakubowska et al., 2006; Li et al., 2002). The presence of a conserved early motif suggests that *sf27* is an early transcribed gene but the expression and function of this ORF have yet to be characterized.

The SfMNPV-3AP2 and SfMNPV-19 genomes both contain a 310 bp deletion with conserved endpoints in the *sf57–sf58* region. This deletion includes the final ten codons of *sf57* (*odv-e66a)*, reducing its length to 691 codons, the complete intergenic region that contains the polyadenylation signals, and the complete unique ORF *sf57a* of SfMNPV-B (Fig. 1). The ortholog of *sf57* in SeMNPV is 723 codons in length. *Sf57* is a paralog of *ac46* from *Autographa californica* MNPV (AcMNPV) that encodes an ODV envelope protein, ODV-E66, that directs reporter proteins to the ODV envelope (Braunagel et al., 2004; Hong et al., 1997). Two copies of *odv-e66* are present in some members of group II NPVs, *odv-e66a* (*sf57*) and *odv-e66b* (*sf115*), whereas just one copy is found in group I NPVs (Ayres et al., 1994; Harrison et al., 2008; Ijkel et al., 1999; Jakubowska et al., 2006; Li et al., 2002); based on sequence identity values, *sf115* was considered to be the ortholog of *ac46*. The biological effects of the presence of two copies of this gene in some NPV genomes have not been studied. This deletion also affects the *sf57a* ORF of unknown function. This ORF is absent in both SfMNPV-3AP2 and SfMNPV-19 (Table S1, Fig. 1). The expression and function of this ORF are currently the subject of study.

Two insertions were found in the SfMNPV-B genome. The longest insertion, 74 bp in length was located in the intergenic region between *sf85* ORF (unique to SfMNPV) and the *sf86* gene (*p26a*), located on opposite strands. Neither the start codon of *sf85*, the promoter, or the stop codon of *sf86* were affected, suggesting that the insertion is unlikely to substantially affect the biological activity of the virus. The second insertion in the SfMNPV-B genome, 61 bp in length, starts 437 nt upstream from the ATG start codon of *sf131*, resulting in a putative product of 299 amino acids, compared to 279 in the other two isolates (Table S1) (Harrison et al., 2008; Wolff et al., 2008) or 278 residues in SeMNPV (Ijkel et al., 1999). *Sf131* is the ortholog of *ac136*, that encodes the P26 protein. *Acp26* is a non-essential gene for AcMNPV replication *in vitro* (Rodems and Friesen, 1993) or *in vivo* (Simón et al., 2008b), and its deletion had no substantial effects on the infectivity of AcMNPV OBs. SfMNPV isolates possess two copies of this gene as do other members of group II NPVs. Deletion of the SfMNPV-B *p26b* gene had no effect on OB infectivity (Simón et al., 2008b).

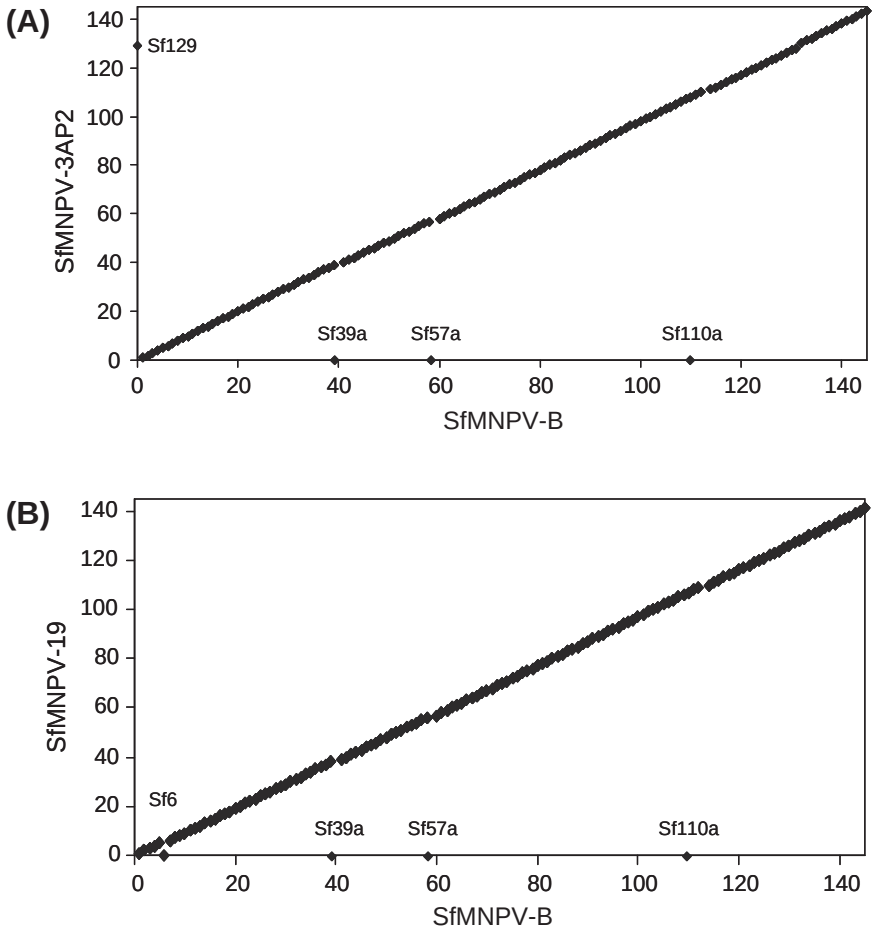

Fig. 1. Gene parity plot comparing ORF content and order of the complete genome of the Nicaraguan isolate of the *Spodoptera frugiperda* MNPV (SfMNPV-B) with the Missouri (SfMNPV-3AP2) and Brazilian (SfMNPV-19) isolates. ORFs present in only one of the compared genomes appear on the axis corresponding to the virus in which they are present.

Within the SfMNPV-19 genome a deletion of 186 bp located in the 5′ region of the *dutpase* gene (*sf53*) was detected when compared with the SfMNPV-B and SfMNPV-3AP2 genomes. In SfMNPV-19 the *dutpase* gene consisted of 124 aa, which is 21 aa shorter than that in SfMNPV-B and SfMNPV-3AP2 (*sf54*) (Table S1). Wolff et al. (2008) indicated that the non-coding region located immediately upstream from the SfMNPV-19 *dutpase* gene was poorly conserved and suggestive of a mutation hotspot.

Finally, a deletion of 192 bp was detected in *hr8* located immediately after the *sf128* gene (*lef-6*) in the SfMNPV-B genome. The effect of this deletion is not known. Both the number and length of *hr* regions are variable among the different baculovirus genomes that have been sequenced to date. The number of *hrs* varies between six in SeMNPV (Ijkel et al., 1999), five in *Agrotis segetum* NPV (AgseNPV) (Jakubowska et al., 2006) and four in MacoNPV (Li et al., 2002). The biological significance of this variability in *hr* number is not known. As *hrs* are possible origins of DNA replication and enhancers of gene expression (Possee and Rohrmann, 1997) this deletion could influence DNA replication and gene expression in SfMNPV-B. A total of eight *hr* regions consisting of one to four imperfect palindromic repeats with conserved repeat size of around 44 nt were detected in the SfMNPV-B genome dispersed along the genome in intergenic regions (Table S1). This pattern concurs with the distribution of *hrs* found in SfMNPV-3AP2 (Harrison et al., 2008) and SfMNPV-19 (Wolff et al., 2008).

Other point mutations and short deletions or insertions that resulted in additional or elimination of codons were identified among the different genomes, the biological implications of which are currently being investigated.

3.2. Comparison of SfMNPV ORFs

The principal differences between isolates at the genome level were due to differences in gene content between genomes. A total of 145 ORFs were identified in the SfMNPV-B genome, defined as methionine-initiated ORFs encoding putative proteins of more than 50 aa and with minimal overlapping with adjacent ORFs (Table S1). Among these ORFs, 141 are common to the three genome sequences. To examine the genome organization, the order of homologous ORFs between all the isolates was compared using gene-parity plot analysis. The gene arrangement and orientation was almost completely collinear between all the isolates, whereas minimal differences were observed in gene content (Fig. 1). SfMNPV-B presented three ORFs (*sf39a*, *sf57a* and *sf110a*) that were not detected in the genomes of SfMNPV-3AP2 (Fig. 1A) and SfMNPV-19 (Fig. 1B). These three ORFs did not overlap adjacent ORFs. The *sf57a* gene encodes a protein comprising 53 aa and is included in the 310 bp deletion present in the SfMNPV-3AP2 and SfMNPV-19 genomes. In contrast, we did not observe major inter-genomic differences (deletions or insertions) at the nucleotide level in the *sf39a* or *sf110a* regions that resulted in the deletion of these genes in SfMNPV-3AP2 or SfMNPV-19. However, both *sf39a* and *sf110a* were located downstream and upstream from homologous regions *hr4* and *hr7*, respectively. Harrison et al. (2008) and

Wolff et al. (2008) considered both regions as *hrs*, whereas in the SfMNPV-B genome two methionine-initiated ORFs larger than 50 aa were identified. In contrast, the SfMNPV-B and SfMNPV-19 genomes lacked the SfMNPV-3AP2 ORF129 homolog. Harrison et al. (2008) identified *sf129* comprising 122 codons between nt 117,291 and 117,659 in *polyhedrin* sense orientation. *Sf130* encodes a 94 aa polypeptide, located between nt 117,372 and 117,656 in anti-*polyhedrin* sense orientation, that completely overlaps the *sf129* ORF. When the *sf129–sf130* region was compared between both genomes, nucleotide sequence variation was observed between nt 117,345 and 117,350 in SfMNPV-3AP2 and nt 118,934 and 118,340 in SfMNPV-B (Fig. 2). This comprises an AAATTT sequence present in the SfMNPV-3AP2 genome, whereas a TATGCGC sequence is present in SfMNPV-B. The latter was followed by four nucleotides and a stop codon four that reduced *sf129* ORF to 21 aa, leading us to reject sf129 as a valid ORF. In addition, *sf129* of SfMNPV-3AP2 has no homologs in other baculovirus genomes, whereas *sf130* had sequence homology with an ORF of unknown function that is present in most baculoviruses; SeMNPV *se128* (Ijkel et al., 1999), *Spodoptera litura* MNPV *splt28* (Pang et al., 2001) or AcMNPV *ac29* (Ayres et al., 1994), suggesting that the *sf129* ORF described by Harrison et al. (2008) may not be a true ORF based on the degree of overlap with *sf130*. Finally, the *sf6* ORF is present in the genomes of SfMNPV-B, SfMNPV-3AP2 and SfMNPV-19, although this appears to have been overlooked in the latter, but no differences in nucleotide sequences were observed among the different genomes in this region. All the other ORFs were conserved between all genomes and showed a high degree of predicted amino acid sequence conservation. In general, nucleotide sequence identities were at least 99.5% with few gaps.

The fall armyworm is a migratory pest that is distributed from Argentina to the United States, with a correspondingly high likelihood of virus dispersal along the insect's migratory routes. This may explain the high similarity found among the different SfMNPV genomes from these areas; high levels of gene conservation were observed between SfMNPV-3AP2 and SfMNPV-19 despite the geographical separation of their points of isolation. SfMNPV-3AP2 purified from a wild-type population is a single genotype with a deletion in the *egt* region. Harrison et al. (2008) also reported the presence of different genotypes within the wild-type population that presented similar deletions to those found in the SfMNPV-NIC isolate (Fig. 3). Genotypes that lacked oral infectivity and *egt* minus genotypes were present in both isolates (Harrison et al., 2008; Simón et al., 2004). It seems that the SfMNPV isolate from Missouri is genotypically structured in a similar way to the SfMNPV-NIC isolate. Indeed, the genotypic structure of the SfMNPV-NIC population appears to maximize the likelihood of transmission (López-Ferber et al., 2003; Simón et al., 2005b, 2008c). Experiments are in progress to determine the similarity at the population level among the different SfMNPVs and its importance in the ecology of the virus.

3.3. Analysis of selection on SfMNPV isolates genes

Selection analysis (Yang, 1998) performed on vertebrate virus genes has identified positively selected sites in regions involved in host immune response and receptor binding (Holmes et al., 2002; Twiddy et al., 2002; Woelk et al., 2001; Woelk and Holmes, 2001). Analysis of variable selection performed among amino acid sites in HIV genes (Yang et al., 2003), *Trypanosoma brucei* essential

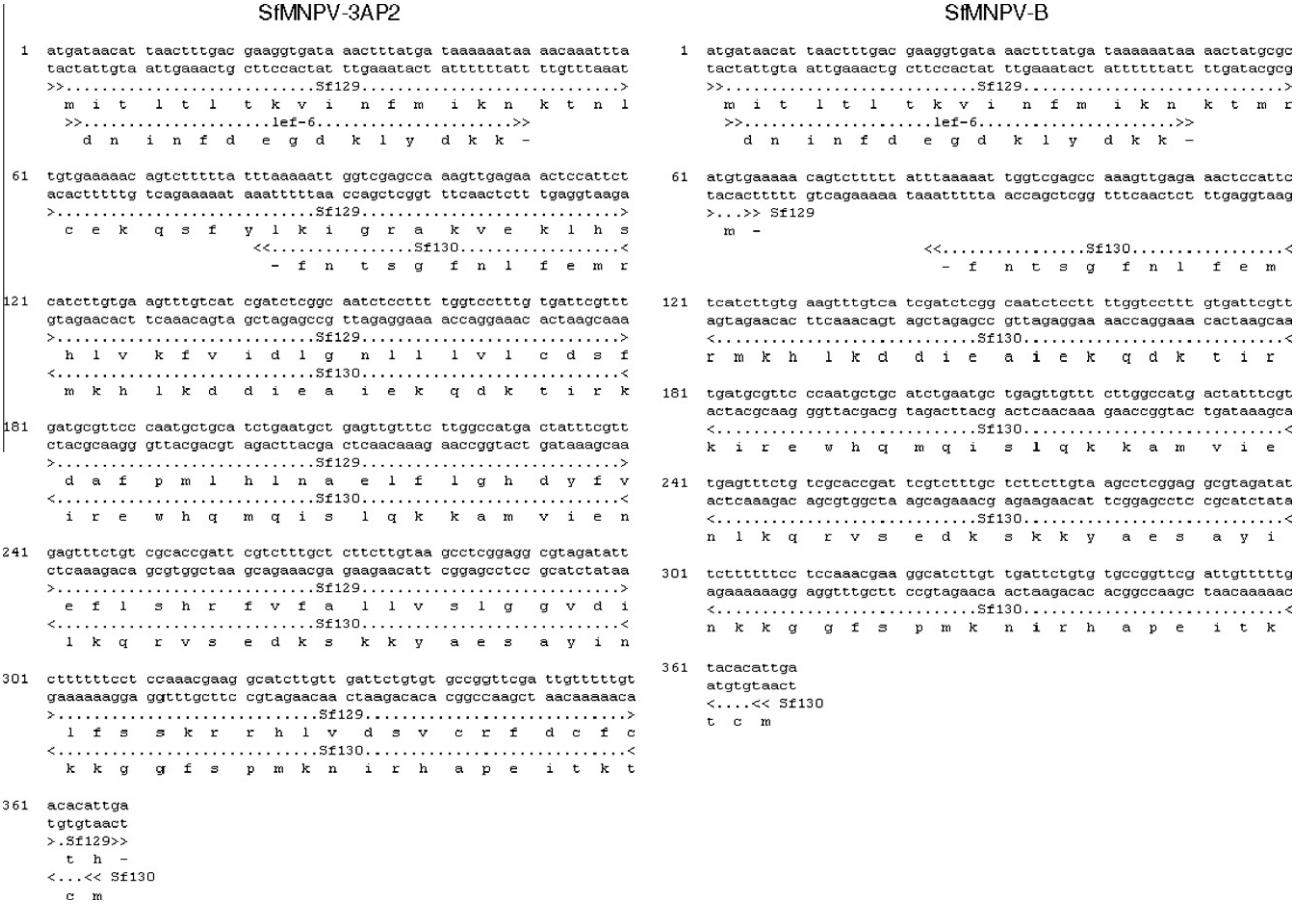

Fig. 2. Sequence comparison between *sf129–sf130* gene regions of SfMNPV-3AP2 and SfMNPV-B genomes.

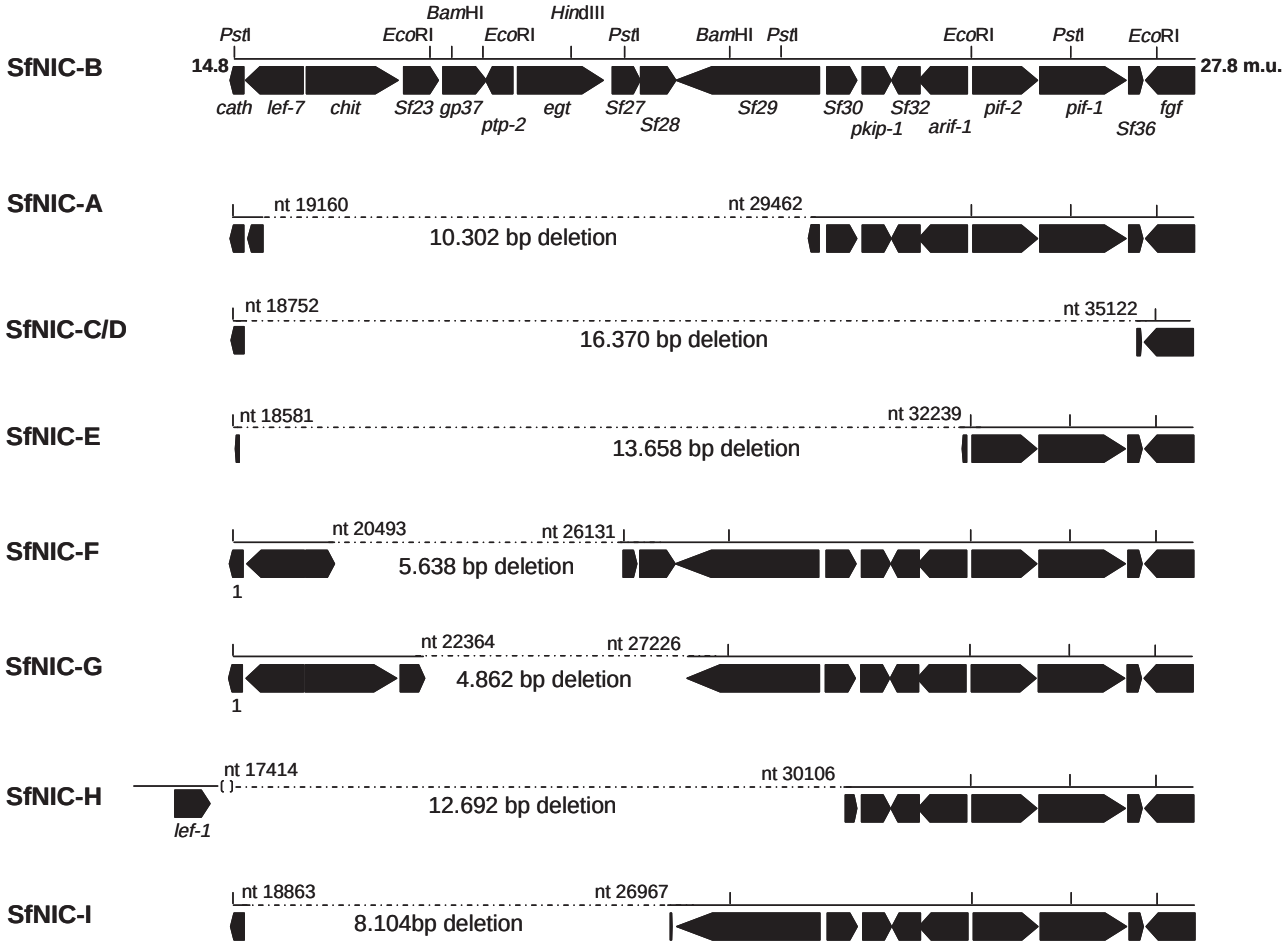

Fig. 3. Schematic representation of the gene order in the genotypes present within the Nicaraguan isolate of SfMNPV. The deletions and the genes absent in each genotype are indicated. The genes and ORFs found in this variable region are indicated bellow the arrows in the complete SfMNPV-B genome that point to their directions of transcription.

surface proteins (Emes and Yang, 2008) and M proteins of influenza A virus (Furuse et al., 2009) revealed sites involved in the adaptive evolution of these intracellular parasites. This technique could help to identify baculovirus genes involved in species-specific virulence and host range, or those involved in adaptation to new host species. Alternatively, positively selected sites could also be involved in selection for improved physical stability or transmission. In the present study, pairwise comparisons and models that allow changes in the value of ω among codon sites were used to evaluate selection on genes from geographically distinct SfMNPV genotypes.

Following the pairwise approach in PAML, the average ω-value was 0.1743 suggesting that negative selection has prevailed during the evolution of most of the ORFs present in SfMNPV genomes (data not shown). When comparing genes among different viruses, overall negative selection values have also been observed (Harrison and Bonning, 2003, 2004). Nevertheless, it is difficult to detect positive selection using a pairwise approach mainly because a large number of amino acids in a protein remains invariant due to strong functional constraints (ω close to 0), requiring that this kind of analysis be complemented with models that allow for heterogeneous ω values among different codon sites (Harrison and Bonning, 2003). In this second analysis, ORF alignments were fitted to six models, M0, M1, M2, M3, M7 and M8 (Harrison and Bonning, 2003, 2004). Models M2, M3 and M8 identified positively selected sites in several ORFs. However, the null hypothesis models (M0, M1 and M7) were rejected only for *sf49*, *sf57* and *sf122*. To rule out sequencing errors that could have generated anomolous results in the selection analysis, these three ORFs were subjected to re-sequencing. For this, *sf49*, *sf57* and *sf122* were amplified from SfMNPV-B DNA with a high fidelity Taq Polymerase (Takara), cloned into pUC19 and re-sequenced using standard primers. The sequences and alignments of the positively selected ORFs from the three SfMNPV genomes showed no evidence of sequencing errors following re-sequencing (Fig. 4). The *sf49* and *sf57* genes have been previously identified as *pif-3* and *odv-e66a*, respectively, whereas no function or protein product has been demonstrated for the *sf122* ORF. Homologs of *sf122* have been found in some, but not all members, of group II NPVs; SeMNPV (*se121*) (Ijkel et al., 1999), *Agrotis ipsilon* NPV (*agip140*) (Harrison, 2008), AgseNPV (*agse133*) (Jakubowska et al., 2006) and SpltNPV (*splt126*) (Pang et al., 2001).

The evidence of positive selection for *pif-3* was stronger than that of *odv-e66a* but weaker than that of *sf122*. Estimation of ω = 0.150 as an average across all sites (M0) suggested that *pif-3* was being subjected to purifying selection (Table 1). However, analyses using models that allow ω to vary among sites, indicated which sites of the *pif-3* gene were subjected to differing selection intensities. Moreover, the likelihood ratio test suggested rejection of the null hypothesis for each of the three comparisons that were performed (Table 2). Models M2, M3 and M8 consistently identified a positively selected site (61C). The ω-values were very high for all three models in approximately 0.5% of the codon sites ($\omega > 150$), indicating a very strong degree of positive selection for

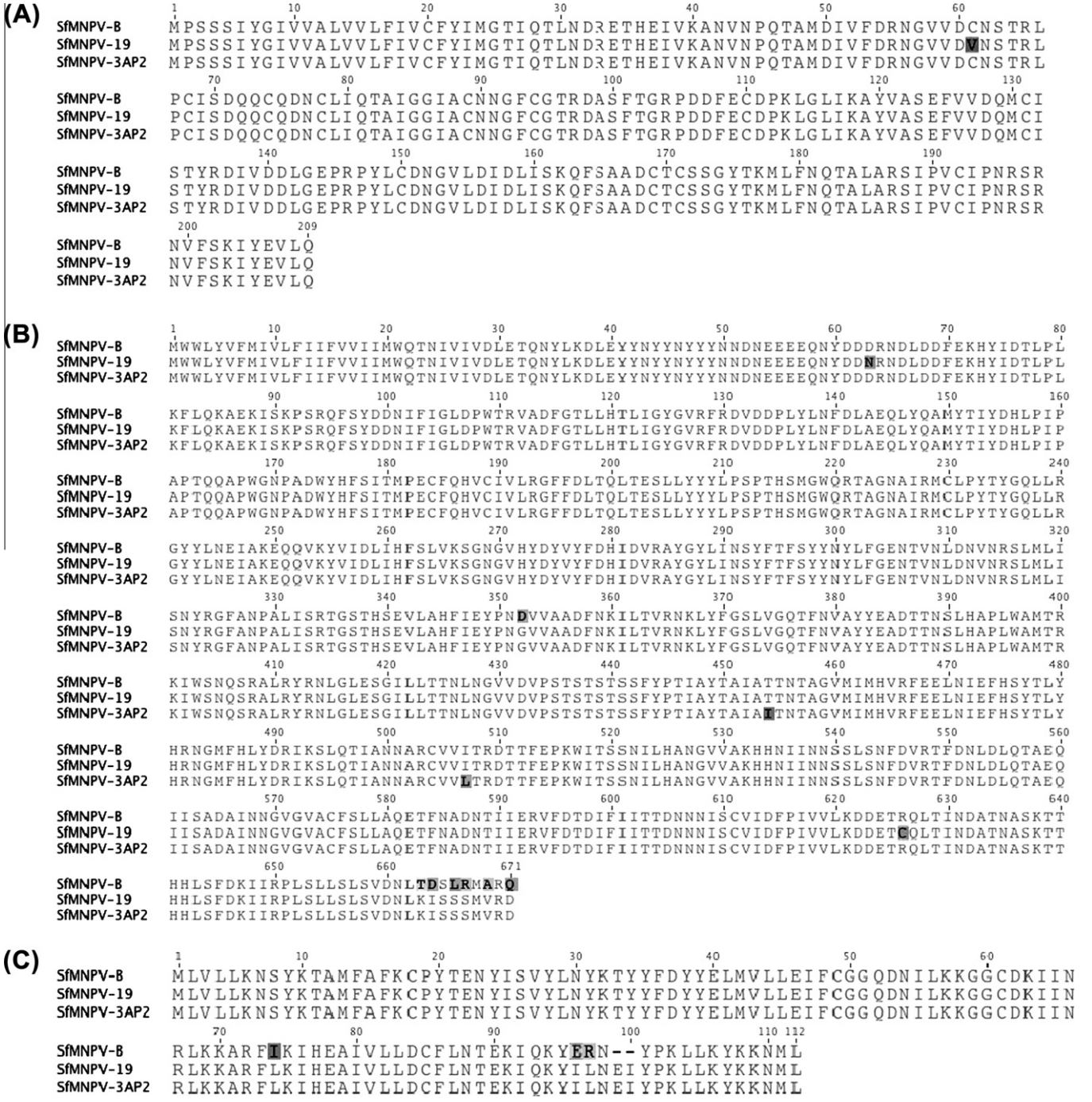

Fig. 4. Multiple sequence alignment of the deduced proteins with positively selected sites from the three SfMNPV genomes; (A) PIF-3, (B) ODV-E66A, and (C) SF122. Amino acid sequences were obtained by translating the nucleotide sequences from the codon alignments. Sites at which positive selection is inferred following removal of indels (BEB analysis) with at least 90% of posterior probabilities are indicated in bold type and gray shading (see Table 3).

this site in *pif-3*. Furthermore, using Bayesian Empirical Bayes (BEB) analysis, site 61 was identified in M2 and M8, as a positively selected site with probabilities greater than 90% but lower than 95% (Table 3).

The *pif-3* gene, along with other ODV envelope protein genes *p74*, *pif-1*, *pif-2*, *pif-4* and *odv-e56* is among the 30 core genes that are found in all baculoviruses sequenced to date (Braunagel and Summers, 2007; Fang et al., 2009; Harrison et al., 2010). These genes encode factors that are essential for *per os* infection (Fang et al., 2006, 2009; Faulkner et al., 1997; Harrison et al., 2010; Kikhno et al., 2002; Pijlman et al., 2003). Peroral infection occurs following a two step process (Horton and Burand, 1993); beginning with ODV binding to specific host cell receptors (Haas-Stapleton et al., 2004; Slack et al., 2008), followed by fusion of the cell and ODV membranes, although the mechanisms by which this occurs remain unclear. Recently, it has been shown that PIF-1, PIF-2 and PIF-3, but not P74 are essential for the formation of a stable complex on the AcMNPV ODV surface, and that P74 is associated with this complex that appears to have an essential function in virus entry (Peng et al., 2010). PIF-3 could therefore be influential in determining host range or infectivity at the level of midgut cell binding and internalization of ODVs.

Table 1
Selection analysis of SfMNPV-B ORF49, SfMNPV-B ORF57 and SfMNPV-B ORF122 and the corresponding ORFs in SfMNPV-3AP2 and SfMNPV-19.

ORF (gene)	M0 (one ratio)	M1 (neutral)	M2 (selection)	M3 (discrete)	M7 (β)	M8 (β and ω)
Sf49 (*pif-3*)	ω = 0.15074	p_0 = 0.94369, ω_0 = 0.000	p_0 = 0.99504, ω_0 = 0.000	p_0 = 0.00001, ω_0 = 0.000	p = 0.00500	p_0 = 0.99504
	lnL: −879.621	p_1 = 0.05631, ω_1 = 1.000	p_1 = 0.00000, ω_1 = 1.000	p_1 = 0.99503, ω_1 = 0.000	q = 0.04990	**p_1 = 0.00496, ω = 150.747**
		lnL: −877.050	**p_2 = 0.00496, ω_2 = 150.745** lnL: −871.537	**p_2 = 0.00496, ω_2 = 150.750** lnL: −871.537	lnL: −877.198	lnL: −871.537
Sf57 (*odv-e66a*)	ω = 0.33816	p_0 = 0.81264, ω_0 = 0.000	p_0 = 0.98427, ω_0 = 0.059	p_0 = 0.97511, ω_0 = 0.00000	p = 0.00500	p_0 = 0.98428
	lnL: −2892.990	p_1 = 0.01873, ω_1 = 1.000	p_1 = 0.00000, ω_1 = 1.000	**p_1 = 0.01799, ω_1 = 15.7487**	q = 0.02028	**p_1 = 0.01572, ω = 38.386**
		lnL: −2886.976	**p_2 = 0.01573, ω_2 = 38.380** lnL: −2875.251	**p_2 = 0.00690, ω_2 = 59.6440** lnL: −2875.237	lnL: −2886.990	lnL: −2875.251
Sf122 (unknown)	ω = 0.15199	p_0 = 0.90199, ω_0 = 0.000	p_0 = 0.98165, ω_0 = 0.07217	p_0 = 0.00000, ω_0 = 0.00000	p = 0.00500	p_0 = 0.9722
	lnL: −492.320	p_1 = 0.09801, ω_1 = 1.000	p_1 = 0.00000, ω_1 = 1.00000	p_1 = 0.09816, ω_1 = 0.07217	q = 0.04780	**p_1 = 0.0277, ω = 999.000**
		lnL: −489.437	**p_2 = 0.01835, ω_2 = 696.701** lnL: −480.945	**p_2 = 0.01835, ω_2 = 696.706** lnL: −480.945	lnL: −489.437	lnL: −481.948

ω ratios greater than 1.0 and corresponding proportions of sites are in bold face.

Table 2
Likelihood ratio statistics ($2\Delta l$) for the comparison of models with variable ω values among sites.

ORF	Probability values		
	M0 vs. M3 (df = 4)	M1 vs. M2 (df = 2)	M7 vs. M8 (df = 2)
sf49 (*pif-3*)	3.019e−03	4.034e−03	3.483e−03
sf57 (*odv-e66a*)	3.656e−07	8.089e−06	7.977e−06
sf122 (unknown)	1.421e−04	2.510e−04	5.592e−04

df: degrees of freedom that correspond to the difference in the numbers of parameters between the two models.

Table 3
Sites identified as evolving under positive selection in models M2 and M8.

ORF	Positively selected sites
sf49 (pif-3)	61C
sf57 (odv-e66a)	663T, **664D**, 666L, **667R**, **671Q**
sf122 (unknown)	**96E**

Sites are numbered according to the reference sequence SfMNPV-B (GenBank accession number HM595733). Positive selection sites were identified using the Bayesian Empirical Bayes (BEB) approach with posterior probability $P \geqslant 90\%$, and with those at $P \geqslant 95\%$ in boldface. The CODEML application was performed following removal of the following indels identified in flanking regions that resulted in frameshifts: *sf49* (*pif-3*): No indels present. *sf57* (*odv-e66a*) Indel 1: TAACAATGACAA deleted between 140 and 141, Indel 2: ATG deleted between 157 and 158, Indel 3: C deleted between 1996 and 1997, Indel 4: TA deleted between 2002 and 2003, Indel 5: variable region with multiple indels and frameshifts (C-terminus removed) ACGTGCCATCCAATTTTTTTACATATGGCGTATGACACGCGTTTCGTGTCATTTTTAAATG ACACGAAACGCGTTTGA deleted between 2013 and 2014. *sf122* Indel 1: A deleted between 294 and 295; Indel 2: GAAATA not deleted, no frameshift between 294 and 295.

For *odv-e66a* the evidence of positive selection was also strong. The average ω value (0.33816) indicates that purifying selection dominates the evolution of the *odv-e66a.* Model M3 (discrete) suggests a proportion of sites (~1.8%) under weaker diversifying selection with ω_1 = 15.7487 and a smaller proportion (~0.7%) under stronger diversifying selection with ω_2 = 59.6440. Under M2 and M8, the estimates suggest that ~1.6% of sites are evolving by positive selection with ω values ~38 (Table 1). All models that allow for positive selection sites reject the null hypothesis in the likelihood ratio tests (Table 2). BEB analysis of models M2 and M8 identified positively selected sites consisting of the same five residues with probabilities greater than 95% (Table 3).

The *odv-e66* is a late gene that encodes for an ODV envelope protein, that redirects protein trafficking to the ODV envelope (Braunagel et al., 2004; Hong et al., 1997). One interesting feature is the presence of two homologs of *odv-e66* (*sf57* and *sf115*) in SfMNPV genomes. *Sf115* appears to be the ortholog of *ac46* characterized as *odv-e66*, whereas *sf57* is the paralog that showed the highest degree of homology to *odv-e66* (AB009613) from *Leucania separate* NPV with 64% identity (ID) and 76% similarity (Sim). Two copies of this gene are also found in other members of group II NPVs such as MacoMNPV, SeMNPV and *Orgyia leucostigma* NPV, whereas a single copy is present in AcMNPV, SpltNPV, AgseNPV or *Helicoverpa armigera* SNPV. The position of these two genes in relation to nearby ORFs is conserved in these NPVs. It is probable that the two copies of the *odv-e66* gene were acquired independently and the first copy originated from a source that was more closely related to LeseNPV and *Lymantria dispar* MNPV than to the ortholog of *ac46* (*sf115*). The presence of two copies of *odv-e66* with both late and early baculovirus promoters may be related to the two forms of *odv-e66* found in mature AcMNPV ODVs (Ijkel et al., 1999; Li et al., 2002). The biological significance of the presence of these two copies in some members of baculoviruses is unknown. In addition, the first copy (*sf57*) is positively selected which is suggestive of diversifying selection, whereas the *sf115* copy did not show this diversify. Such directional selection may cause alterations in the activity of the encoded protein, facilitating the adaptation to a new host species and/or overcoming host defenses.

Finally, analysis of the M2, M3 and M8 models for ORF *sf122* resulted in extremely high ω values that are usually obtained when

d_S tends to zero. The average ω value (~0.152) in M0 indicates relatively strong purifying selection acting on *sf122* (Table 1). M0, M1 and M7 models were rejected by likelihood ratio tests (Table 2). BEB analysis of M2, and M8 identified three sites as being positively selected, but only one (96E) with probabilities greater than 95% (Table 3).

Sf122 encodes a putative protein 139 aa in length of unknown function that is present in some members of group II NPVs. The *sf122* transcripts were classified as late transcribed by RT-PCR (data not shown), suggesting that this is a functional late gene. BLAST searches revealed homologies to SeMNPV SE121 (49% ID and 69% Sim), SpltNPV SPLT126 (43% ID and 67% Sim), AgseNPV AGSE133 (38% ID and 61% Sim) and AgipNPV AGIP140 (33% ID and 62% Sim). Sequence comparison revealed that the SF122 protein presented local areas of homology in different regions along the proteins of other organisms, suggesting that those regions could be related to specific protein domains. SF122 is a small putative protein of around 16.7 kDa, without conserved domains. SF122 TargetP analyses demonstrated that the amino acid sequence contained a cytoplasmic membrane localization signal supported by a reliability class value of RC = 3. SF122 SignalP analysis using gram-negative and gram-positive networks, identified this protein as a non-secreted protein. Moreover, virus-Ploc indicated the inner layer of the cytoplasmic membrane as the likely localization site of the putative protein. Further studies will be performed to determine the function of this protein in SfMNPV.

In summary, the likelihood ratio test rejected the null model in each of the three comparisons for the three positively selected genes: M0 (one ratio) against M3 (discrete), M1 (neutral) against M2 (selection), and M7 (β) against M8 (β and ω) (Table 2). This suggests diversifying selection for *pif-3*, *odv-e66a* and *sf122*. The *pif-3* and *odv-e66a* genes encode proteins involved in primary infection, whereas no function has yet been described for the late transcribed gene. The primary infection is the first barrier to diversification for these viruses. Diversifying selection may favor alterations in the activity of the encoded protein, facilitating the adaptation to a new host species or overcoming the immune response in the novel host. Experiments are in progress to determine the possible function of these three proteins, particularly the unknown protein SF122. The three geographical SfMNPV genotypes differ significantly in a number of biological activities that have direct relevance to their use as biological insecticides (Harrison et al., 2008; Wolff et al., 2008; Simón et al., 2004).

The ability of the empirical method to detect sites is defined by the probability of rejecting the null hypothesis (neutral selection) when it is wrong, leading us to accept that the alternative hypothesis, positive selection, is likely correct. This probability decreases in accuracy when a low number of sequences are analyzed (Anisimova et al., 2001). It may be expected that from the analysis of only three genome sequences of the same virus species, as reported here, we were only able to reject the null hypothesis models in three instances for which the strength of selection was very high. Similar results were obtained when comparing two closely related sequences (AcMNPV, RoMNPV), for which positively selected sites were detected in only two genes (Harrison and Bonning, 2003). Selection analysis that includes sequences from other SfMNPV isolates or even from other NPV genomes may help to identify a greater diversity of viral genes that are subjected to positive selection.

Acknowledgments

We thank N. Gorria (Universidad Pública de Navarra, Pamplona, Spain) for technical assistance. This work was supported by the Grant AGL200507909-C03-01. TW acknowledges support from CONACYT infrastructure Grant I0-110/162/10.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jip.2011.01.002.

References

Altschul, C.H., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 229, 381–399.
Anisimova, M., Bielawski, J.P., Yang, Z., 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol. Biol. Evol. 18, 1585–1592.
Argaud, O., Croizier, L., López-Ferber, M., Croizier, G., 1998. Two key mutations in the host-range specificity domain of the *p143* gene of *Autographa californica* nucleopolyhedrovirus are required to kill *Bombyx mori* larvae. J. Gen. Virol. 79, 931–935.
Ayres, M.D., Howard, S.C., Kuzio, J., Lopez-Ferber, M., Possee, R.D., 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virology 202, 586–605.
Barreto, M.E., Guimaraes, C.T., Teixeira, F.F., Paiva, E., Valicente, F.H., 2005. Effect of *Baculovirus spodoptera* isolates in *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae) larvae and their characterization by RAPD. Neotrop. Entomol. 34, 208–282.
Bendtsen, J.D., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: signalP 3.0. J. Mol. Biol. 340, 783–795.
Braunagel, S.C., Summers, M.D., 2007. Molecular biology of the baculovirus occlusion-derived virus envelope. Curr. Drug Targets 8, 1084–1095.
Braunagel, S.C., Williamson, S.T., Saksena, S., Zhong, Z., Russell, W., Russell, D.H., Summers, M.D., 2004. Trafficking of ODV-E66 is mediated via a sorting motif and other viral proteins: facilitated trafficking to the inner nuclear membrane. Proc. Natl. Acad. Sci. USA 22, 8372–8377.
Chen, C.J., Thiem, S.M., 1997. Differential infectivity of two *Autographa californica* nucleopolyhedrovirus mutants on three permissive cell lines is the result of *lef-7* deletion. Virology 227, 88–95.
Chen, C.J., Quentin, M.E., Brennan, L.A., Kukel, C., Thiem, S.M., 1998. *Lymantria dispar* nucleopolyhedrovirus *hrf-1* expands the larval host range of *Autographa californica* nucleopolyhedrovirus. J. Virol. 72, 2526–2531.
Chen, X., Sun, X., Hu, Z., Li, M., ÓReilly, D.R., Zuidema, D., Vlak, J.M., 2000. Genetic engineering of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus as an improved pesticide. J. Invertebr. Pathol. 76, 140–146.
Chou, K.C., Shen, H.B., 2008. Cell-PLoc: a package of web-servers for predicting subcellular localization of proteins in various organisms. Nat. Protoc. 3, 153–162.
Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016.
Emes, R.D., Yang, Z., 2008. Duplicated paralogous genes subject to positive selection in the genome of *Trypanosoma brucei*. PLoS One 3, e2295.
Erlandson, M., 2009. Genetic variation in field populations of baculoviruses: mechanisms for generating variation and its potential role in baculovirus epizootiology. Virol. Sin. 24, 458–469.
Erlandson, M., Newhouse, S., Moore, K., Janmaat, A., Myers, J., Theilmann, D., 2007. Characterization of baculovirus isolates from *Trichoplusia ni* populations from vegetable greenhouses. Biol. Control 41, 256–263.
Escribano, A., Williams, T., Goulson, D., Cave, R.D., Chapman, J.W., Caballero, P., 1999. Selection of a nucleopolyhedrovirus for control of *Spodoptera frugiperda* (Lepidoptera: Noctuidae): structural, genetic and biological comparison of four isolates from the Americas. J. Econ. Entomol. 92, 1079–1085.
Fang, M.G., Nie, Y.C., Wang, Q., Deng, F., Wang, R.R., Wang, H.Z., Wang, H.L., Vlak, J.M., Chen, X.W., Hu, Z.H., 2006. Open reading frame 132 of *Helicoverpa armigera* nucleopolyhedrovirus encodes a functional *per os* infectivity factor (PIF-2). J. Gen. Virol. 87, 2563–2569.
Fang, M., Nie, Y., Harris, S., Erlandson, M.A., Theilmann, D.A., 2009. *Autographa californica* multiple nucleopolyhedrovirus core gene *ac96* encodes a *per os* infectivity factor (*pif-4*). J. Virol. 83, 12569–12578.
Faulkner, P., Kuzio, J., Williams, G.V., Wilson, J.A., 1997. Analysis of P74, a PDV envelope protein of *Autographa californica* nucleopolyhedrovirus required for occlusion body infectivity in vivo. J. Gen. Virol. 78, 3091–3100.
Frengen, E., Weichenhan, D., Zhao, B., Osoegawa, K., van Geel, M., de Jong, P.J., 1999. A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. Genomics 58, 250–253.
Furuse, Y., Suzuki, A., Kamigaki, T., Oshitani, H., 2009. Evolution of the M gene of the influenza A virus in different host species: large scale sequence analysis. Virol. J. 6, 1–13.
Fuxa, J.R., 1982. Prevalence of viral-infections in populations of fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in southeastern Louisiana. Environ. Entomol. 11, 239–242.
Haas-Stapleton, E.J., Washburn, J.O., Volkman, L.E., 2004. P74 mediates specific binding of *Autographa californica* M nucleopolyhedrovirus occlusion-derived virus to primary cellular targets in the midgut epithelia of *Heliothis virescens* larvae. J. Virol. 78, 6786–6791.
Harrison, R.L., 2008. Genomic sequence analysis of the Illinois strain of the *Agrotis ipsilon* multiple nucleopolyhedrovirus. Virus Genes 38, 155–170.

Harrison, R.L., Bonning, B.C., 2003. Comparative analysis of the genomes of *Rachiplusia ou* and *Autographa californica* multiple nucleopolyhedrovirus. J. Gen. Virol. 84, 1827–1842.

Harrison, R.L., Bonning, B.C., 2004. Application of maximum-likelihood models to selection pressure analysis of group I nucleopolyhedrovirus genes. J. Gen. Virol. 85, 197–210.

Harrison, R.L., Puttler, B., Popham, H.J.R., 2008. Genomic sequence analysis of a fast-killing isolate of *Spodoptera frugiperda* multiple nucleopolyhedrovirus. J. Gen. Virol. 89, 775–790.

Harrison, R.L., Sparks, W.O., Bonning, B.C., 2010. *Autographa californica* multiple nucleopolyhedrovirus ODV-E56 envelope protein is required for oral infectivity and can be substituted functionally by *Rachiplusia ou* multiple nucleopolyhedrovirus ODV-E56. J. Gen. Virol. 91, 1173–1182.

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

Hong, T., Summers, M.D., Braunagel, S.C., 1997. N-terminal sequences from *Autographa californica* nuclear polyhedrosis virus envelope proteins ODV-E66 and ODV-E25 are sufficient to direct reporter proteins to the nuclear envelope, intranuclear microvesicles and the envelope of occlusion derived virus. Proc. Natl. Acad. Sci. USA 94, 4050–4055.

Horton, H.M., Burand, J.P., 1993. Saturable attachment sites for polyhedron-derived baculovirus on insect cells and evidence for entry via direct membrane fusion. J. Virol. 67, 1860–1868.

IJkel, W.F., van Strien, E.A., Heldens, J.G., Broer, R., Zuidema, D., Goldbach, R.W., Vlak, J.M., 1999. Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. J. Gen. Virol. 80, 3013–3022.

Jakubowska, A.K., Peters, S.A., Ziemnicka, J., Vlak, J.M., van Oers, M.M., 2006. Genome sequence of an enhancin gene-rich nucleopolyhedrovirus (NPV) from *Agrotis segetum*: colinearity with *Spodoptera exigua* multiple NPV. J. Gen. Virol. 87, 537–551.

Kamita, S.G., Maeda, S., 1997. Sequencing of the putative DNA helicase-enconding gene of the *Bombyx mori* nuclear polyhedrosis virus and fine-mapping of a region involved in host range expansion. Gene 190, 173–179.

Kikhno, I., Gutiérrez, S., Croizier, L., Croizier, G., López-Ferber, M., 2002. Characterization of *pif*, a gene required for the *per os* infectivity of *Spodoptera littoralis* nucleopolyhedrovirus. J. Gen. Virol. 83, 3013–3022.

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

López-Ferber, M., Simon, O., Williams, T., Caballero, P., 2003. Defective or effective? Mutualistic interactions between virus genotypes. Proc. Biol. Sci. 270, 2249–2255.

ÓReilly, D.R., 1995. Baculovirus-encoded ecdysteroid UDP-glucosyltransferases. Insect Biochem. Mol. Biol. 25, 541–550.

ÓReilly, D.R., Miller, L.K., 1991. Improvement of a baculovirus pesticide by deletion of the *egt* gene. Biotechnology 9, 1086–1089.

Pang, Y., Yu, J., Wang, L., Hu, X., Li, W., Chen, C., Han, H., Hu, S., Yang, H., 2001. Sequence analysis of the *Spodoptera litura* multicapsid nucleopolyhedrovirus genome. Virology 287, 391–404.

Pearson, W.R., 1990. Rapid and sensitive comparison with FASTP and FASTA. Methods Enzymol. 183, 63–98.

Peng, K., van Oers, M.M., Hu, Z., van Lent, J.W.M., Vlak, J.M., 2010. Baculovirus *per os* infectivity factors form a complex on the surface of occlusion-derived virus. J. Virol. 84, 9497–9504.

Pijlman, G.P., Pruijssers, A.J., Vlak, J.M., 2003. Identification of *pif-2*, a third conserved baculovirus gene required for *per os* infection of insects. J. Gen. Virol. 84, 2041–2049.

Popham, H.J.R., Pellock, B.J., Robson, M., Dierks, P.M., Miller, L.K., 1998. Characterization of a variant of *Autographa californica* nuclear polyhedrosis virus with a non-functional ORF 603. Biol. Control 12, 223–230.

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

Rodems, S.M., Friesen, P.D., 1993. The *hr5* transcriptional enhancer stimulates early expression from *Autographa californica* nuclear polyhedrosis virus genome but is not required for virus replication. J. Virol. 67, 5776–5785.

Shapiro, D.I., Fuxa, J.R., Braymer, H.D., Pashley, D.P., 1991. DNA restriction polymorphism in wild isolates of *Spodoptera frugiperda* nuclear polyhedrosis virus. J. Invertebr. Pathol. 58, 96–105.

Simón, O., Williams, T., Lopez-Ferber, M., Caballero, P., 2004. Genetic structure of a *Spodoptera frugiperda* nucleopolyhedrovirus population: high prevalence of deletion genotypes. Appl. Environ. Microbiol. 70, 5579–5588.

Simón, O., Chevent, F., Williams, T., Caballero, P., Lopez-Ferber, M., 2005a. Physical and partial genetic map of *Spodoptera frugiperda* nucleopolyhedrovirus (SfMNPV) genome. Virus Genes 30, 403–417.

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

Simón, O., Williams, T., Asensio, A.C., Ros, S., Gaya, A., Caballero, P., Possee, R.D., 2008a. *Sf29* gene of *Spodoptera frugiperda* multiple nucleopolyhedrovirus is a viral factor that determines the number of virions in occlusion bodies. J. Virol. 82, 7897–7904.

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

Simón, O., Williams, T., López-Ferber, M., Taulemesse, J.M., Caballero, P., 2008c. Population genetic structure determines speed of kill and occlusion body production in *Spodoptera frugiperda* multiple nucleopolyhedrovirus. Biol. Control 44, 321–330.

Simón, O., Williams, T., Possee, R.D., López-Ferber, M., Caballero, P., 2010. Stability of a *Spodoptera frugiperda* nucleopolyhedrovirus deletion mutant during serial passages in insects. Appl. Environ. Microbiol. 76, 803–809.

Slack, J.M., Lawrence, S.D., Krell, P.J., Arif, B.M., 2008. Trypsin cleavage of the baculovirus occlusion-derived virus attachment protein P74 is prerequisite in *per os* infection. J. Gen. Virol. 89, 2388–2397.

Slavicek, J.M., Popham, H.J.R., Reigel, C.I., 1999. Deletion of the *Lymantria dispar* multicapsid nucleopolyhedrovirus ecdysteroid UDP-glucosyl transferase gene enhances viral killing speed in the last instar of the gypsy moth. Biol. Control 16, 91–103.

Treacy, M.F., All, J.N., Ghidiu, G.M., 1997. Effect of ecdysteroid UDP-glucosyltransferase gene deletion in efficacy of a baculovirus against *Heliothis virescens* and *Trichoplusia ni* (Lepidoptera: Noctuidae). J. Econ. Entomol. 90, 1207–1214.

Tumilasci, V.F., Leal, E., Zanotto, P.M., Luque, T., Wolff, J.L., 2003. Sequence analysis of a 5.1 kbp region of the *Spodoptera frugiperda* multicapsid nucleopolyhedrovirus genome that comprises a functional ecdysteroid UDP-glucosyltransferase (*egt*) gene. Virus Genes 27, 137–144.

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

Williams, T., Goulson, D., Caballero, P., Cisneros, J., Martinez, A.M., Chapman, J.W., Roman, D.X., Cave, R.D., 1999. Evaluation of a baculovirus bioinsecticide for small-scale maize growers in Latin America. Biol. Control 14, 67–75.

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

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

Wolff, J.L., Valicente, F.H., Martins, R., Oliveira, J.V., Zanotto, P.M., 2008. Analysis of the genome of *Spodoptera frugiperda* nucleopolyhedrovirus (SfMNPV-19) and of the high genomic heterogeneity in group II nucleopolyhedroviruses. J. Gen. Virol. 89, 1202–1211.

Yang, Z., 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15, 568–573.

Yang, Z., 2007. Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 15, 1586–1591.

Yang, W., Bielawsky, J.P., Yang, Z., 2003. Widespread adaptative evolution in the Human immunodeficiency virus type 1 genome. J. Mol. Evol. 57, 212–221.