



HAL
open science

Transcription of densovirus endogenous sequences in the *Myzus persicae* genome

Gabriel Clavijo, Manuella van Munster, Baptiste Monsion, Nicole Bochet,
Véronique Brault

► **To cite this version:**

Gabriel Clavijo, Manuella van Munster, Baptiste Monsion, Nicole Bochet, Véronique Brault. Transcription of densovirus endogenous sequences in the *Myzus persicae* genome. *Journal of General Virology*, 2016, 97, pp.1000-1009. 10.1099/jgv.0.000396 . hal-02636488

HAL Id: hal-02636488

<https://hal.inrae.fr/hal-02636488>

Submitted on 29 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Transcription of densovirus endogenous sequences in the *Myzus persicae* genome

Gabriel Clavijo,^{1†} Manuella van Munster,^{2†} Baptiste Monsion,^{1‡} Nicole Bochet¹ and Véronique Brault¹

¹UMR 1131 SVQV INRA-UDS, 28 rue de Herrlisheim, 68021 Colmar, France

²UMR 385 BGPI, INRA-CIRAD-SupAgro, CIRAD TA-A54/K, Campus International de Baillarguet, 34398 Montpellier, France

Correspondence

Véronique Brault
veronique.brault@colmar.inra.fr

Integration of non-retroviral sequences in the genome of different organisms has been observed and, in some cases, a relationship of these integrations with immunity has been established. The genome of the green peach aphid, *Myzus persicae* (clone G006), was screened for densovirus-like sequence (DLS) integrations. A total of 21 DLSs localized on 10 scaffolds were retrieved that mostly shared sequence identity with two aphid-infecting viruses, *Myzus persicae* densovirus (MpDNV) and *Dysaphis plantaginea* densovirus (DpDNV). In some cases, uninterrupted potential ORFs corresponding to non-structural viral proteins or capsid proteins were found within DLSs identified in the aphid genome. In particular, one scaffold harboured a complete virus-like genome, while another scaffold contained two virus-like genomes in reverse orientation. Remarkably, transcription of some of these ORFs was observed in *M. persicae*, suggesting a biological effect of these viral integrations. In contrast to most of the other densoviruses identified so far that induce acute host infection, it has been reported previously that MpDNV has only a minor effect on *M. persicae* fitness, while DpDNV can even have a beneficial effect on its aphid host. This suggests that DLS integration in the *M. persicae* genome may be responsible for the latency of MpDNV infection in the aphid host.

Received 23 November 2015

Accepted 8 January 2016

INTRODUCTION

Insects are the most diverse and largest animal group, causing economically significant losses in a wide range of crops, either as direct pests or as disease vectors (Chilana *et al.*, 2012). Studying insect biology and virus transmission by insects is therefore critical to identify control targets to minimize their impact. Advances in sequencing technology have dramatically increased the amount of information available on insect genomes, creating new opportunities for development of research programmes focused on insect biology. Deep-sequence analysis of insect genomes has revealed that they may contain integrated viral sequences that are not of retroviral origin (Katzourakis & Gifford, 2010; Liu *et al.*, 2011b; Thézé *et al.*, 2011). These ancestral endogenous sequences share similarity with viruses in different families: ssDNA viruses (family *Parvoviridae*), dsDNA viruses (families *Baculoviridae*,

Polydnaviridae and *Nudiviridae*), reverse-transcribed DNA viruses (family *Hepadnaviridae*), ssRNA positive-sense (family *Flaviviridae*) or negative-sense (families *Rhabdoviridae* and *Orthomyxoviridae*) viruses, and dsRNA viruses (family *Reoviridae*) (Bézier *et al.*, 2009; Cheng *et al.*, 2014; Katzourakis & Gifford, 2010; Liu *et al.*, 2011c; Thézé *et al.*, 2011). Integration of such a great diversity of viral sequences suggests that many different replication strategies can induce integration of these sequences into host genomes. Typically, viral sequences are found scattered throughout insect genomes, and in most cases, no complete virus genome has been found. However, in some cases, integrated viral ORFs are uninterrupted and can therefore potentially be expressed in host cells (Bézier *et al.*, 2009; Katzourakis & Gifford, 2010; Liu *et al.*, 2011b). In plants, insects and vertebrates, integration of viral sequences is predicted to confer selective advantages to the host by generating antiviral immunity against related viruses (Belyi *et al.*, 2010; Bertsch *et al.*, 2009; Flegel, 2009; Goic *et al.*, 2013; Maori *et al.*, 2007).

Viral sequences from viruses in the subfamily *Densovirinae* (family *Parvoviridae*) have been shown to be integrated into the genome of *Drosophila*, *Rhodnius prolixus* and the pea aphid (*Acyrtosiphon pisum*) (Liu *et al.*, 2011b). Densoviruses (DNVs) are small, non-enveloped ssDNA

[†]These authors contributed equally to this study.

[‡]Present address: Institut de Biologie Moléculaire des Plantes, Laboratoire propre du CNRS conventionné avec l'Université de Strasbourg, 12 rue du Général Zimmer, 67084 Strasbourg, France.

One supplementary figure and five tables are available with the online Supplementary Material.

viruses encapsidated in icosahedral particles of about 25 nm diameter. Their genomes are 4–6 kb and encode only a few proteins referred to as either non-structural (NS) or capsid (CP, also referred to as VP) proteins (Bergoin & Tijssen, 2010). NS proteins are implicated in virus replication and in the activation of the promoter driving the synthesis of CP proteins (Legendre & Rommelaere, 1992; Rhode & Richard, 1987), while CP proteins are involved in virion formation (Croizier *et al.*, 2000). Genome organization varies among the different genera in the subfamily, with the two sets of genes being in tandem on the same DNA strand in the genera *Iteravirus*, *Brevidensovirus*, *Hepandensovirus* and *Penstyl-densovirus* and on complementary strands in the genus *Ambidensovirus*. DNVs are found in a wide range of insect families and can cause acute diseases that are often fatal for their insect host (Barreau *et al.*, 1996; Mutuel *et al.*, 2010; Suto, 1979; Szelei *et al.*, 2011). As DNVs are not known to infect mammals (Dupont, 2003; El-Far *et al.*, 2004), biopesticides derived from engineered or natural DNVs have been developed to reduce lepidopteran, mosquito and cockroach populations (Buchatskii *et al.*, 1987; Carlson *et al.*, 2006; Hu *et al.*, 2000; Jiang *et al.*, 2007; Monsarrat *et al.*, 1984; Tal & Attathom, 1993). However, some DNVs may only induce sublethal infection and may even confer a beneficial effect on their host (i) by increasing the production of the winged morph and therefore promoting insect dispersal, or (ii) by protecting against pathogenic baculovirus infection (Ryabov *et al.*, 2009; van Munster *et al.*, 2003a; Xu *et al.*, 2014).

We utilized the recent release of the genome sequence of the aphid *Myzus persicae* to look for integrated viral sequences originating from viruses of the subfamily *Densovirinae*. We identified and confirmed the presence of genomic sequences sharing partial homology with *Myzus persicae* densovirus (MpDNV) and *Dysaphis plantaginea* densovirus (DplDNV) (Ryabov *et al.*, 2009; van Munster *et al.*, 2003b). Interestingly, we also showed that these integrated viral sequences are transcribed in *M. persicae*. Previous studies have shown that MpDNV has a moderate effect on the development and fecundity of *M. persicae* with no significant effect on the intrinsic growth rate of the population (van Munster *et al.*, 2003a). It is possible that integration and transcription of densovirus-like sequences (DLSs) sharing similarities with MpDNV could be responsible for the low pathogenicity of MpDNV in *M. persicae*.

RESULTS

Detection and organization of DLSs in the *M. persicae* genome

We screened the assembled genome of *M. persicae* (clone G006) for integrated sequences homologous to DNV, hereafter referred to as DLSs. BLAST searches were performed using as queries the reference strain sequences from each

of the five genera in the subfamily *Densovirinae* (*Brevidensovirus*, *Hepandensovirus*, *Iteradensovirus*, *Penstyl-densovirus* and *Ambidensovirus*) together with the genome sequences of the two known aphid DNVs, MpDNV and DplDNV. Twenty-one genomic sequences sharing significant sequence identity with MpDNV and DplDNV (E-values ranging from 2.10^{-116} to 1.10^{-4}) were identified on 10 scaffolds (Table 1). An additional genomic sequence from scaffold_883 exhibited sequence identity with the lepidopteran densovirus *Junonia coenia* densovirus (JcDNV), the type member of the genus *Ambidensovirus* (Table 1). DLS integrations covered genomic sequence lengths ranging from 87 to 1887 bp. Embedded in these sequences, deduced amino acid sequences of different lengths (from 29 to 664 aa) shared 24–74 % identity with NS or CP proteins of MpDNV, DplDNV or JcDNV (Table 1).

An almost-complete ambidensovirus-like genome was found on scaffold_1839 with the NS protein-coding sequences on one strand of the genome and the CP protein-coding sequences on the complementary strand. This organization of the coding sequences was consistent with that of genomes found in infectious virions (Fig. 1). A more peculiar situation was encountered on scaffold_794 where two almost-complete ambidensovirus-like genomes were integrated, both in reverse orientation (Fig. 1). This coding sequence structure may imply two independent but neighbouring insertions in the genome. The other scaffolds contained ambidensovirus-like sequences sharing homology with either NS or CP protein sequences. In most cases, these viral sequences were only partial. As several integrations were detected at the beginning or at the end of a scaffold (scaffold_564, _1826, _1646, _1268 and _1885), we cannot completely exclude the existence of additional virus-like sequences that may be missing in the available data.

The presence of DLS integrations in *M. persicae* genome was confirmed in another available *M. persicae* genome sequence (clone_O). All the integration events detected in the G006 clone were observed in clone_O (Table S1, available in the online Supplementary Material). This analysis was further extended to the genome of another aphid species, *A. pisum* (International Aphid Genomics Consortium, 2010). Using an updated full-genome sequence of *A. pisum* (assembly v.2), we retrieved 53 DLS integrations (Table S2). Previously, Liu *et al.* (2011b) described the presence of 49 DLS integrations in the *A. pisum* genome. The discrepancy between our analyses and those of Liu *et al.* (2011b) might result from the new assembly of an *A. pisum* genome. In contrast to the *M. persicae* genome, the DLS integrations in *A. pisum* shared sequence identity not only with MpDNV, DplDNV and JcDNV but also with two other DNVs belonging to the genus *Ambidensovirus*, *Galleria mellonella* densovirus (GmDNV) and *Helicoverpa armigera* densovirus (HaDNV) (Table S2). The observed DLS integrations covered genomic sequence lengths ranging from 165 to 462 bp. Portions of these sequences (55–1547 aa) shared amino acid sequence identity ranging from 23 to 46% when compared with the

Table 1. Inventory of DLS integrations within the *M. persicae* genome (clone G006) using five genera of DNVs as queries

| Scaffold no. | Scaffold size (bp) | Nt position of integration | DNV genus | Top viral BLAST hit (reference strain) | Related gene | GenBank accession no. | Aa ID (%) | Alignment length (aa) | E-value | Name of DLS in Phylogeny |
|---------------|--------------------|----------------------------|-----------------------|--|--------------|-----------------------------|-----------|-----------------------|---------|--------------------------|
| scaffold_28 | 1 079 708 | 674429–675373 | <i>Ambidensovirus</i> | <i>Dysaphis plantaginea</i> densovirus | NS | gi204324151 gb ACI01073_11 | 43.0 | 316 | 2E–76 | MpDLS_28 |
| scaffold_28 | 1 079 708 | 673450–673863 | <i>Ambidensovirus</i> | <i>Dysaphis plantaginea</i> densovirus | NS | gi204324151 gb ACI01073_11 | 24.0 | 138 | 1E–08 | |
| scaffold_413 | 266 173 | 193157–192810 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | CP | gi33235701 ref NP_874378_11 | 42.2 | 116 | 2E–25 | |
| scaffold_564 | 193 318 | 1124–1231 | <i>Ambidensovirus</i> | <i>Dysaphis plantaginea</i> densovirus | CP | gi204324153 gb ACI01075_11 | 58.3 | 36 | 1E–04 | |
| scaffold_794 | 108 655 | 11516–12106 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | CP | gi33235702 ref NP_874377_11 | 33.0 | 200 | 5E–24 | |
| scaffold_794 | 108 655 | 13426–12422 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | NS | gi33235700 ref NP_874376_11 | 51.2 | 365 | 3E–99 | MpDLS_794-1 |
| scaffold_794 | 108 655 | 13904–13491 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | NS | gi33235699 ref NP_874375_11 | 36.4 | 143 | 3E–14 | |
| scaffold_794 | 108 655 | 15297–15734 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | NS | gi33235699 ref NP_874375_11 | 35.8 | 151 | 1E–14 | |
| scaffold_794 | 108 655 | 15799–16803 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | NS | gi33235700 ref NP_874376_11 | 51.2 | 365 | 3E–99 | MpDLS_794-2 |
| scaffold_794 | 108 655 | 18741–17119 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | CP | gi33235702 ref NP_874377_11 | 41.1 | 560 | 1E–108 | |
| scaffold_794 | 108 655 | 18750–18601 | <i>Ambidensovirus</i> | <i>Dysaphis plantaginea</i> densovirus | CP | gi204324153 gb ACI01075_11 | 74.00 | 50 | 5E–18 | |
| scaffold_794 | 108 655 | 19148–18762 | <i>Ambidensovirus</i> | <i>Dysaphis plantaginea</i> densovirus | CP | gi204324154 gb ACI01076_11 | 41.9 | 136 | 2E–20 | |
| scaffold_883 | 69 518 | 62012–62188 | <i>Ambidensovirus</i> | <i>Junonia coenia</i> densovirus | CP | gi257676 gb AAB23698_11 | 39.0 | 59 | 3E–04 | |
| scaffold_1268 | 27 708 | 714–370 | <i>Ambidensovirus</i> | <i>Dysaphis plantaginea</i> densovirus | CP | gi204324154 gb ACI01076_11 | 60.9 | 115 | 1E–34 | |
| scaffold_1335 | 26 965 | 15334–15789 | <i>Ambidensovirus</i> | <i>Dysaphis plantaginea</i> densovirus | CP | gi204324154 gb ACI01076_11 | 40.3 | 154 | 1E–22 | |
| scaffold_1646 | 16 671 | 298–212 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | NS | gi33235699 ref NP_874375_11 | 51.7 | 29 | 8E–08 | |
| scaffold_1646 | 16 671 | 456–316 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | NS | gi33235699 ref NP_874375_11 | 34.7 | 49 | 8E–08 | |
| scaffold_1826 | 10 113 | 2042–156 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | NS | gi33235700 ref NP_874376_11 | 37.5 | 664 | 5E–105 | MpDLS_1826 |
| scaffold_1839 | 9 957 | 6335–7213 | <i>Ambidensovirus</i> | <i>Dysaphis plantaginea</i> densovirus | NS | gi204324151 gb ACI01073_11 | 63.0 | 292 | 2E–116 | MpDLS_1839 |
| scaffold_1839 | 9 957 | 9111–7447 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | CP | gi33235702 ref NP_874377_11 | 40.6 | 571 | 7E–105 | |
| scaffold_1839 | 9 957 | 9505–9041 | <i>Ambidensovirus</i> | <i>Myzus persicae</i> densovirus | CP | gi33235701 ref NP_874378_11 | 33.7 | 166 | 2E–13 | |
| scaffold_1885 | 8 960 | 8151–8591 | <i>Ambidensovirus</i> | <i>Dysaphis plantaginea</i> densovirus | CP | gi204324154 gb ACI01076_11 | 51.0 | 149 | 4E–38 | |

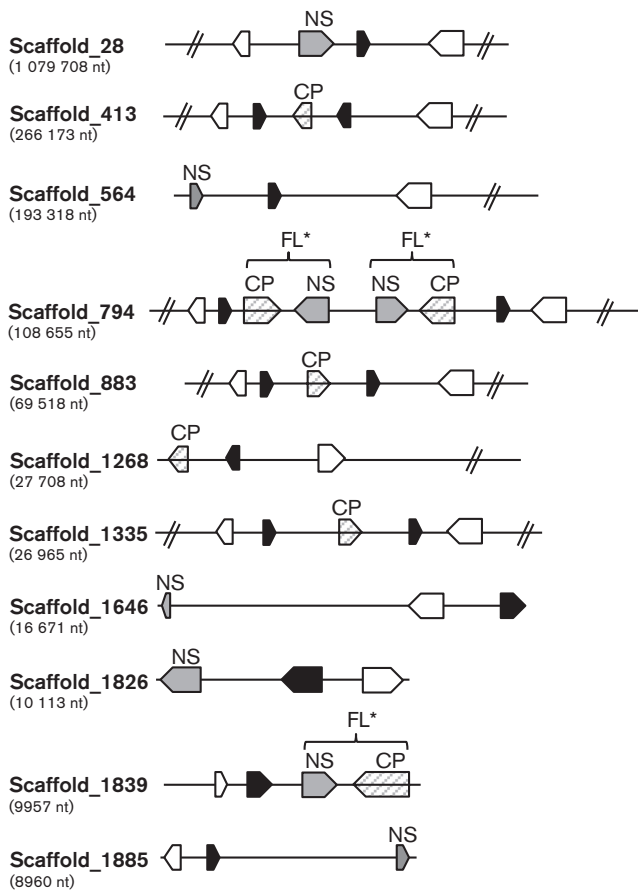


Fig. 1. DLSs and surrounding transposable elements (TEs) in the *M. persicae* genome. The shaded arrows correspond to DLSs with similarity to NS proteins of DNV, the hatched grey arrows to DLSs with similarity to CP proteins of DNVs, the filled arrows to TEs and the open arrows to host genes. The size of each scaffold is indicated in parentheses. The position of full-length DLSs is shown as FL*.

reference sequences of MpDENV, DpDENV, JcDENV, GmDENV and HaDENV proteins (Table S2).

The probability that these endogenous viral sequences correspond to DLSs was reinforced by the prediction of functional ORFs corresponding to either CP or NS proteins (Fig. 2, Table S3). On scaffold_1839, three ORFs potentially encoded one NS and two CP proteins, whereas on scaffold_794 each densovirus full-length integration could encode two NS and one CP protein (Fig. 2, Table S3). DLS integrations were confirmed experimentally by PCR amplification on an *M. persicae* clone originating from Colmar (France). In order to avoid potential non-specific detection of non-integrated MpDENV sequences, free MpDENV was eliminated from this clone by isolating individual larvae immediately after birth and depositing them on sugar beets (see Methods). A first set of primers was designed to specifically amplify the internal integrated sequences showing the highest nucleotide identity with the

aphid DNV sequences (primers referred to as ‘integration’ in Table S4; primer positions shown in Fig. 2a). As shown in Fig. 2(b, upper panel), the presence of the integrated viral sequences was confirmed by amplification from the virus-free aphid genome of DNA fragments of the expected sizes (Table S4). The pair of primers for P_{XV} amplified two fragments on the *M. persicae* genome: one of the expected size (142 bp) and a second representing a longer fragment (~200 bp), probably corresponding either to a non-specific hybridization in the aphid genome (Fig. 2b, upper panel) or to a larger sequence that is fragmented in the current genome assembly upon which fragment size predictions were based. A second set of primers was designed to amplify DNA fragments encompassing insect and viral-like sequences (primers referred to as ‘junction’ in Table S4; primer positions shown in Fig. 2a). These fragments of larger size (1054–4131 bp) were also amplified from the *M. persicae* genome, although with a lower efficiency when using the two pairs of primers for P_I and P_{IX} (Fig. 2b, lower panel). Taken together, these results confirmed the presence of DLSs in the genome of *M. persicae*.

Phylogenetic analysis

To evaluate the genetic relationships between DLSs present in both the *M. persicae* and *A. pisum* genomes with known DNV sequences, we performed a phylogenetic analysis using all integrated sequences that contained a 300 bp DNA sequence from the NS1 gene. This sequence corresponds to highly conserved domains (DNA helicase and ATPase domains) of the *rep* gene in parvoviruses. As the nucleotide sequences were too divergent to be reliably aligned, we used the translated sequence of all 300 bp sequences (100 aa) to build a phylogenetic tree. Two nodes potentially corresponding to different integration events were observed (Fig. 3). The first cluster contained a mixture of DLSs found in the *M. persicae* and *A. pisum* genomes that grouped with MpDENV and DpDENV (Hub 1, bootstrap value 93%). In particular, *M. persicae* DLSs seemed to be closely related to these two aphid DNVs (bootstrap value 82%). The second cluster was almost exclusively composed of DLSs from *A. pisum*, and might result from a different integration event (Fig. 3). Only one DLS in the *M. persicae* genome (MpDLS_28) grouped in this cluster, but the tree robustness at this level (Hub 2, bootstrap value 15%) was not sufficient to confirm with confidence the affiliation of MpDLS_28 to this cluster.

Transposable element (TE) detection in proximity to DLS integrations

We addressed the question of the frequency of TEs in proximity to DLS integrations to identify elements possibly responsible for the integration of the infectious virus sequence into the host genome (endogenization). Some parvoviruses appear to have evolved towards a transposon-like propagation during evolution, as was shown in

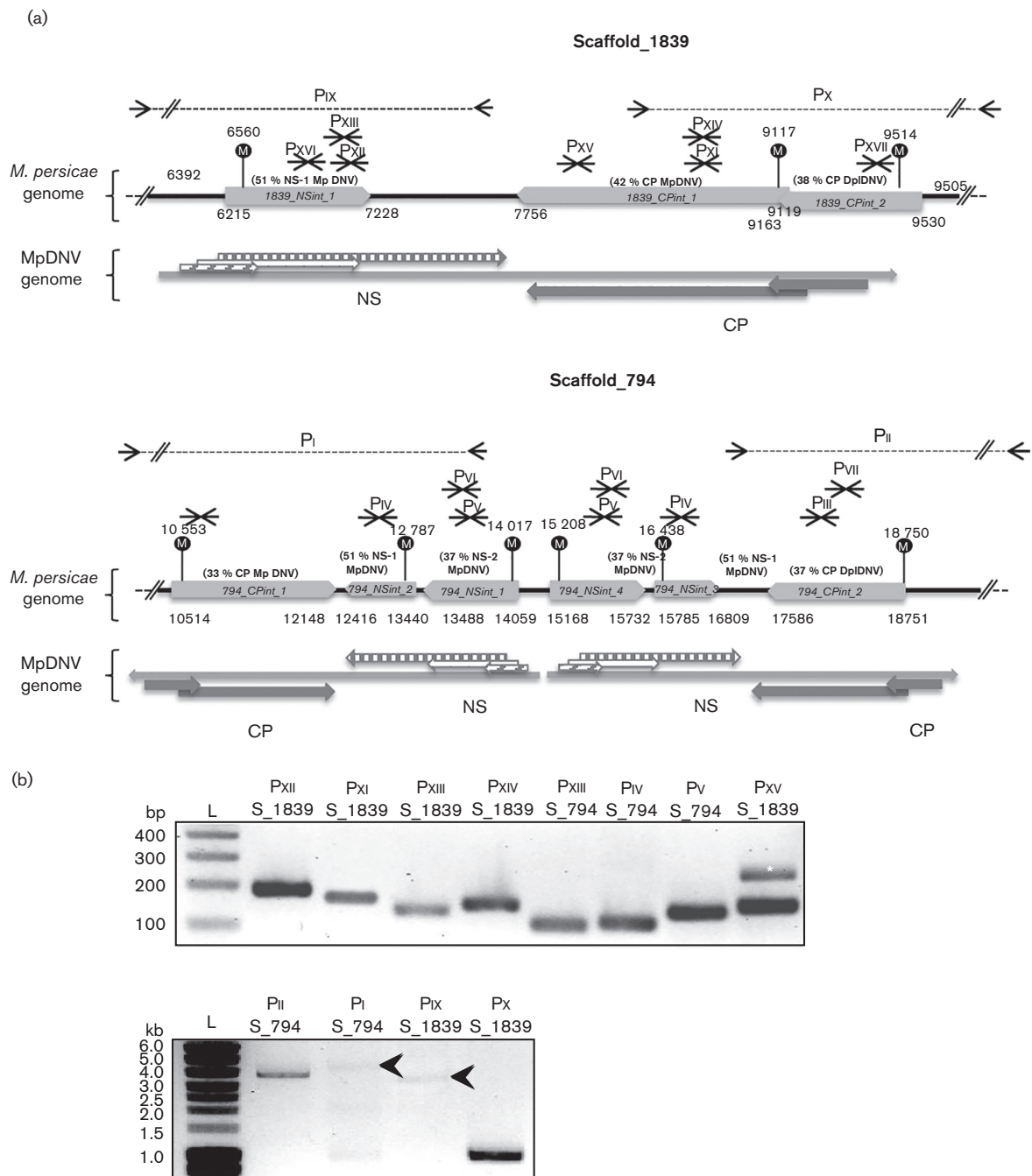


Fig. 2. (a) Schematic representation of DLSs in scaffold_1839 (upper panel) and scaffold_794 (lower panel) of the *M. persicae* genome (clone G006). Shaded arrows on the black line stand for potential ORFs of DLS (*NSint* and *CPint*). The percentage identity with the corresponding sequence in the genome of MpDNV is indicated above. The nucleotide position of these potential ORFs in the scaffold together with the first methionine ([INSERT Fig. 002][INSERT Fig. 001]) in the sequence is indicated. The MpDNV genome organization is presented below with variously shaded arrowheads. The position of primers used to confirm DLS integrations and transcription is shown by short arrows. (b) Confirmation of DLS integration in scaffold_1839 and scaffold_794 in the genome of *M. persicae* DNA (clone Colmar) by PCR amplification of internal integrated viral sequences (upper panel) or junctions between DLSs and the *M. persicae* sequence (lower panel). PCR products were separated by gel electrophoresis on a 2% (upper panel) or 0.7% (lower panel) agarose gel. The names of primers and the targeted scaffolds (S) are indicated above each lane. The position of molecular mass markers (lane L) is shown on the left. The asterisk indicates a non-specific amplification, while arrowheads show DNA fragments of low intensity. The primer sequences are listed in Table S4.

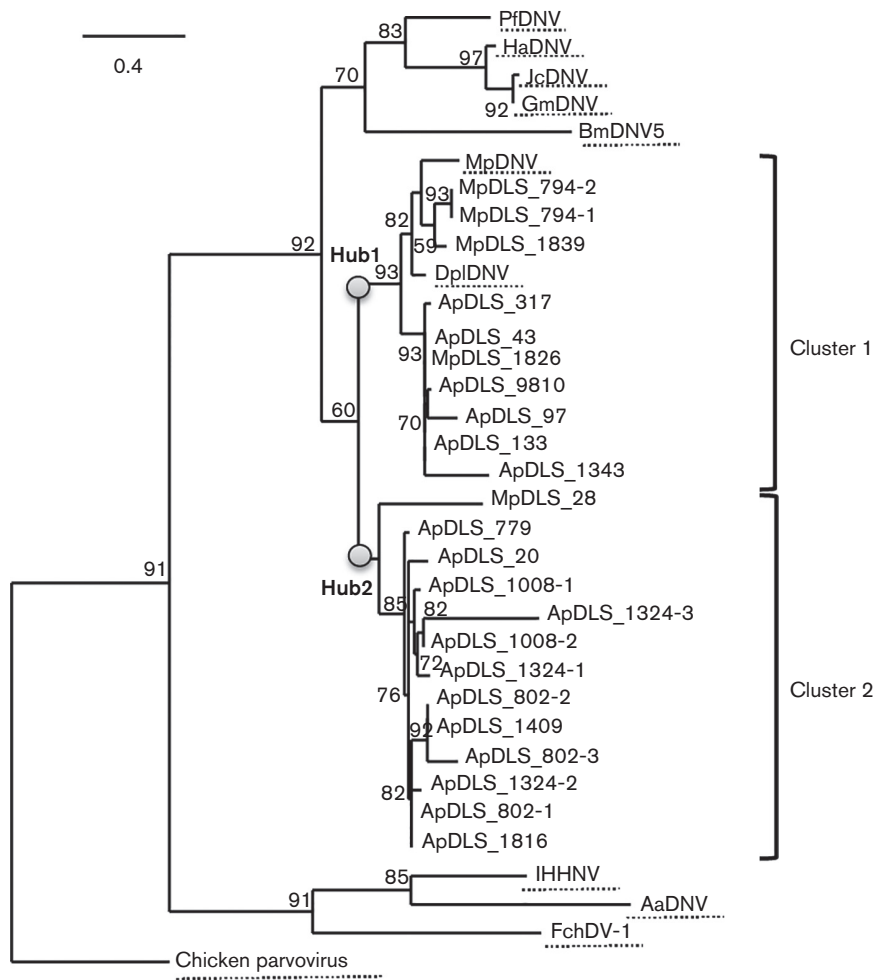


Fig. 3. Phylogenetic relationships of DLS integrations in *M. persicae*, *A. pisum* and exogenous DNVs (underlined). DLS nomenclature is consistent with Tables 1 and S2. Bootstrap values correspond to 1000 replications and are indicated for values >50%. Sequences of 100 aa corresponding to translation of the integrated DLSs in *M. persicae* (see Table 1) and *A. pisum* genomes (see Table S2) sharing similarity with NS1 of parvoviruses were aligned with densovirus sequences arising from: *Aedes albopictus* densovirus (AaDNV; GenBank accession no. NC_004285.1), infectious hypodermal and haematopoietic necrosis virus (IHHNV; NC_002190.2), *Bombyx mori* densovirus 5 (BmDNV5; NC_004287.1), *Fenneropenaeus chinensis* hepatopancreatic densovirus (FchDV-1; NC_014357.1), JcDNV (NC_004284.1), MpDNV (NC_005040.1), DplDNV (EU851411.1), *Periplaneta fuliginosa* densovirus (PfdNV; NC_000936.1), GmDNV (NC_004286.1) and HaDNV (JQ894784.1). Chicken parvovirus NS1 (AJB28748.1) was used to root the tree. The scale represents probabilities of change from one amino acid to another in terms of a unit, which is an expected 1% change between two amino acid sequences.

the acorn worm *Saccoglossus kowalevskii* (Liu *et al.*, 2011a). We investigated the presence of TEs in genomic sequences of the clone G006 of *M. persicae* covering 30 kb upstream and downstream of each DLS using Rebase and National Center for Biotechnology Information (NCBI) sequence analysis tools (BLASTX and BLASTN) (Table S5). In four scaffolds (scaffold_794, _413, _1335 and _883), we detected TEs on both sides of DLS integrations (Fig. 1). On the other scaffolds, only one TE was located at either the 3' or 5' side of the DLS (Fig. 1). Several full-length

(or almost-complete) sequences of TEs were present in some scaffolds (Table S5). Most of the putative TEs were DNA transposons from class II in the classification originally proposed by Finnegan (1989). Despite the presence of TEs in scaffolds containing DLSS, their presence could not be correlated with the presence of DLSS, as a similar frequency of TE integrations was also observed in DLS-free regions (data not shown). We also did not observe terminal inverted repeats flanking the DLSS, which are hallmarks of proliferation by transposition (data not shown).

Transcription of DLSs in *M. persicae*

We addressed whether DLSs in the *M. persicae* genome were transcribed by performing reverse transcription (RT)-PCR and real-time RT-PCR experiments on polyadenylated mRNA extracted from whole aphids (clone Colmar deprived from MpDNV) (see Fig. 2 for position of the potential DLS ORFs). Absence of residual genomic DNA was first confirmed by a lack of amplification when the PCRs were performed directly on total RNA (data not shown). Primers (Table S4) were designed to specifically analyse transcription of each of the potential viral coding sequences on scaffold_1839 and _794. The expected DNA fragments were obtained for the NSint_1 and CPint_2 ORFs on scaffold_1839 or from the CPint_1 and CPint_2 ORFs on scaffold_794 (data not shown). Amplification was also observed using the primer pair for P_{IV} (Fig. 2), which suggested a potential transcription of NSint_1 and/or NSint_4. Indeed, the high sequence identity between these two ORFs prevented detection of each specific mRNA. The specificity of the amplification reactions was confirmed by sequencing the amplified fragments. In contrast, no fragments or fragments of unexpected size were obtained using the other sets of primers amplifying potential transcripts from CPint_1 on scaffold_1839 or NSint_2 and/or

NSint_3 on scaffold_794, suggesting an absence of transcription of these potential ORFs (data not shown). In a second step, in order to measure the rate of accumulation of the viral transcripts detected by RT-PCR, four independent real-time RT-PCRs were performed on polyadenylated mRNAs extracted from whole aphids, and the results of one representative experiment is presented in Fig. 4. The relative accumulation of mRNA transcribed from DLS ORFs was normalized to two housekeeping genes (RLP7 and L27). Interestingly, we observed amplification of mRNA for each of the DLS integrations assayed (Fig. 4). These expressions were confirmed in three additional experiments, although some variation in the transcription levels of CPint_1 on scaffold_794 and of CPint_2 on scaffold_1839 was observed between the different biological replicates (Fig. S1). Despite these variations, the tendency was in favour of a higher transcription of the DNV CP-like genes compared with the DNV NS-like genes.

DISCUSSION

Our study provides evidence of the presence of DLSs in the *M. persicae* genome. Integration of viral genomes belonging to the subfamily *Densovirinae* (family *Parvoviridae*) into arthropods has been reported previously and, in particular, DLSs have been identified in the genome of the pea aphid, *A. pisum* (Liu *et al.*, 2011b). The presence of DLSs in the genome of *M. persicae* implies that the ancestor of *M. persicae* was infected with a DNV whose genome was endogenized. Three almost-complete ambidensovirus-like genomes were integrated into the genome of the two sequenced *M. persicae* clones, which have distinct geographical origins (clone_O from Europe and clone G006 from the USA). This suggests old and stable integration events. The high capacity of parvoviruses to integrate into their host genome is probably due to the virus replication cycle in the nucleus, which requires synthesis of a double-stranded replicative form and the presence of inverted terminal repeats that facilitate integration into the host DNA (Berns, 1990; Yang *et al.*, 1997). The mechanism by which DLS integrations occurred in the *M. persicae* genome is still unknown but is not thought to be linked to TEs, as no hallmarks of transposition were found in the proximities of DLSs. It is therefore likely that the integration paradigm of adeno-associated virus, which has been extensively studied, could apply to other members of the family *Parvoviridae* including DNVs. Adeno-associated virus integration relies on the endonuclease activity of its rolling-circle replication protein, a function that is shared with the NS1 protein of DNV (Ding *et al.*, 2002; Krupovic & Forterre, 2015).

Interestingly, potential coding capacities retained within the DLSs were identified, and transcription in *M. persicae* of the integrated NS- or CP-encoding sequences (hereafter referred to as NSint and CPint) was confirmed. As the predicted viral ORFs were not all transcribed, and the transcript levels from transcribed ORFs appeared unequal,

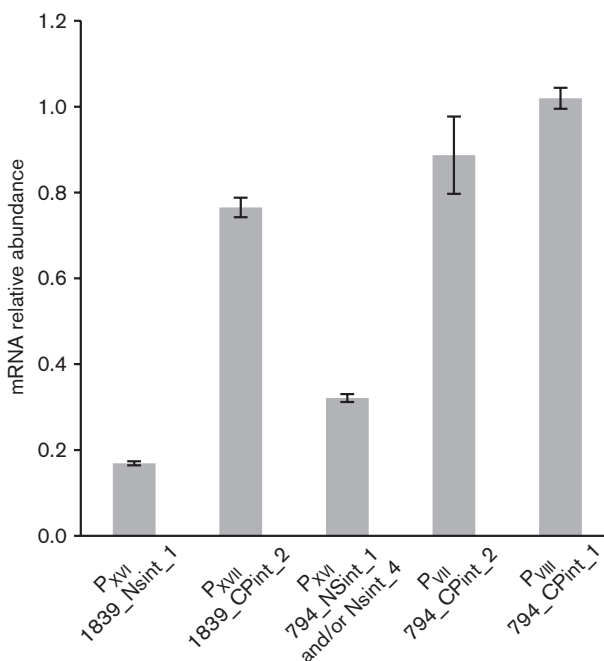


Fig. 4. Transcription profiles of DLSs in *M. persicae*. The relative mRNA abundance of potential DLS ORFs in *M. persicae* (clone Colmar without free MpDNV) were measured by real-time RT-PCR and normalized to the values for the ribosomal housekeeping genes RLP7 and L27. Error bars correspond to technical triplicates. The names of primers (see Table S4) and potential ORFs transcribed are indicated below each bar.

the presence of functional promoters was assessed (Neural Network Promoter Prediction software, http://www.fruitfly.org/seq_tools/promoter.html; data not shown). In scaffold_1839, potential promoters upstream of CPint_2 were identified, whereas no promoter sequences were detected upstream of CPint_1. This observation could explain the lack of CPint_1 transcription observed in our experiments. On the same scaffold, potential promoters that could control transcription of NSint_1 were also found. On the other scaffold (scaffold_794), promoter-like sequences with the potential to control transcription of CPint_1 and CPint_2 were found upstream of each ORF. Transcription of NSint_1 and/or NSint_4 could also be explained by the presence of promoter-like sequences upstream of these two ORFs. As occurs in MpDENV, NS ORFs are supposed to be transcribed from a single promoter (van Munster *et al.*, 2003b), so transcription of the downstream NSint_2 and NSint_3 was expected but not observed in our experiments. The absence of transcription of these ORFs could indicate a transcription termination signal between the two NSint ORFs. Altogether, promoter sequence identification coincided with our detection of viral transcripts. All potential promoter sequences driving transcription of the endogenous viral ORFs are probably of aphid origin, as they do not share any sequence similarity with the predicted promoters in MpDENV or with other viral sequences of MpDENV or DpDENV (data not shown). The higher level of transcription of CP-like genes compared with NS-like genes could therefore reflect a difference in endogenous host promoter transcription efficiency. Transcription of endogenized parvovirus sequences in *A. pisum* was also suggested previously, as expressed sequence tags exhibiting similarities to parvovirus-related DNA were reported in the NCBI public databases (Liu *et al.*, 2011b). However, no experimental confirmation of mRNA accumulation of these parvovirus-related DNAs was reported. Although we clearly detected polyadenylated RNA originating from DLSs in *M. persicae*, the question of their translation remains.

Transcription, and potentially translation, of these sequences could explain the low pathogenicity of DNV in aphids. To date, there are only two DNVs described in aphids: MpDENV, which is highly specific to *M. persicae* (van Munster *et al.*, 2003a), and DpDENV, identified in *D. plantaginea* (Ryabov *et al.*, 2009). Interestingly, these aphid DNVs do not harm their host, in contrast to other DNVs which are highly pathogenic to their insect hosts (Barreau *et al.*, 1996; Liu *et al.*, 2011c; Mutuel *et al.*, 2010; Suto, 1979). MpDENV and DpDENV only slightly affect the development or reproduction of aphids (Ryabov *et al.*, 2009; van Munster *et al.*, 2003a). Moreover, the existence of a selective advantage for the aphid to harbour the virus has been shown for DpDENV, which confers a beneficial effect on *D. plantaginea* by inducing the production of winged morphs at a low insect density, therefore promoting aphid dispersal (Ryabov *et al.*,

2009). Concerning MpDENV, although there is a clear advantage for the virus to maintain a suboptimal infection that does not kill the host (van Munster *et al.*, 2003a), no beneficial effect of MpDENV on *M. persicae* has been discovered so far (V. Brault, unpublished).

The cause for latency in DNV infections of these viruses within aphid hosts could be due to DLS integrations in the aphid genome, as reported in this study. Indeed, transcription of these viral integrated sequences could prime a silencing-based defence mechanism that could target cognate viruses and inhibit the development of an acute infection, as has been shown for *Drosophila* (Goic *et al.*, 2013) and bees (Maori *et al.*, 2007). In this respect, DLSs in scaffold_794 or _1839 contained 12 stretches of 21–29 nt sharing either a perfect nucleotide sequence identity with MpDENV or having only one or two mismatches with MpDENV or DpDENV (Table S6). Although these sequence homologies were limited to a small portion of the viral genome, they were of an appropriate length to induce viral RNA degradation, or translation inhibition, which could control and limit infectious DNV multiplication. Another mechanism of protection induced by the integrated viral sequences could be the synthesis of non-functional DNV-like proteins that could inhibit virus multiplication, as has been shown for human immunodeficiency virus type 1 (HIV-1). Expression of a Gag protein from a human endogenous retrovirus impaired progeny release and HIV-1 infectivity (Monde *et al.*, 2012). The formation of empty particles from endogenized CP protein genes able to prime insect immune protection against further infections could also be an alternative hypothesis explaining protection conferred by viral endogenized sequences. In addition, the activation of the immune system of the insect could be responsible for reducing the pathogenicity of the DNV. These different aspects of virus control should be addressed in the future.

METHODS

Insect rearing and virus clearing. An *M. persicae* (Sulzer) clone from Colmar (France) was reared on pepper (*Capsicum annuum*) at 20 °C with a 16 h photoperiod. This clone was infected with MpDENV. The same clone was cured of MpDENV by transferring individual larvae directly after birth onto beetles (*Beta vulgaris*) to avoid contamination by contact. The absence of MpDENV was monitored by PCR (van Munster *et al.*, 2003a), and one aphid colony cured of MpDENV was maintained on *B. vulgaris*.

DLS detection in the *M. persicae* and *A. pisum* genomes. To screen for DLSs in the *M. persicae* genome, BLASTN and tBLASTn searches were performed online against the AphidBase database (<http://www.aphidbase.com/aphidbase/>), using *M. persicae* clone G006 (originating from the USA; Ramsey *et al.*, 2007) as a reference genome (scaffolds). Full DNA genomes of the following subfamily *Densovirinae* member types: infectious hypodermal and haematopoietic necrosis virus (IHNV, genus *Penstylidensovirus*), *Aedes albopictus* densovirus (AaDENV, genus *Breviendensovirus*), *Bombyx mori* densovirus 5 (BmDENV5, genus *Iteradensovirus*), *Fenneropenaenus chinensis* hepadensovirus (FchDV-1, genus *Hepadensovirus*) and

JcDNV (genus *Ambidensovirus*), as well as full DNA genomes and NS and CP protein sequences of MpDNV and DplDNV, were used as queries. All non-redundant hits from these searches with E-values of $<1.10^{-3}$ were extracted and used in further analyses.

To eliminate possible chimeras or errors in DLS regions of the genome assembly, BLAST searches using extracted DLSs from *M. persicae* clone G006 were performed against the genome of *M. persicae* clone_O (available in AphidBase and originating from the UK). In a similar manner, BLAST searches were performed on the updated assembly of the *A. pisum* genome (assembly v.2) also available in AphidBase. *A. pisum* sequences showing similarity to subfamily *Densovirinae* representatives were also identified and extracted for phylogenetic analysis.

Integration of DLSs in *M. persicae* clones was confirmed by amplifying virus-like sequences using specific primer pairs designed with the Primer3 program (Rozen & Skaletsky, 2000) and with binding sites in both flanking host and virus insertion sequence (Table S4, Fig. 2). Total genomic DNA was extracted from 30 aphids, which were ground in liquid nitrogen using the QIAamp DNA Mini kit protocol (Qiagen). One microlitre of DNA ($20\text{--}30\text{ ng }\mu\text{l}^{-1}$) was used for PCRs containing the specific primers and performed in a thermal cycler (Eppendorf) using standard protocols. PCR products were analysed by 0.7 or 2% agarose gel electrophoresis depending on amplicon size and visualized under UV after ethidium bromide staining.

Alignment and phylogenetic reconstruction. Phylogenetic trees were reconstructed using the predicted amino acid sequences of DLSs. Protein sequences of 100 aa corresponding to a region of the DLS sharing similarity with NS1 of parvoviruses were aligned with the corresponding sequences of known DNVs including subtypes of the subfamily *Densovirinae* using the MUSCLE program configured for high accuracy (Edgar, 2004). The DNV sequences used in the alignment were: AaDNV (GenBank accession no. NC_004285.1), IHNV (NC_002190.2), BmDNV5 (NC_004287.1), FchDV-1 (NC_014357.1), JcDNV, (NC_004284.1), MpDNV (NC_005040.1), DplDNV (EU851411.1), *Periplaneta fuliginosa* densovirus (NC_000936.1), GmDNV (NC_004286.1) and HaDNV (JQ894784.1). Phylogenetic relationships among DLS integrations and DNVs were assessed using the Phylogeny.fr platform (Dereeper *et al.*, 2008). The phylogenetic tree was reconstructed using the maximum-likelihood method implemented in the PhyML program (v.3.1/3.0 aLRT). The JTT 1 G model (Jones *et al.*, 1992), a commonly used probabilistic amino acid replacement matrix at the 85% identity level allowing heterogeneity across sites, was selected. Reliability for the internal branch was assessed using the aLRT test (SH-Like). Node support was estimated using bootstrap resampling procedures (1000 trials). Trees were edited using TREE-DYN (v.198.3) (Chevenet *et al.*, 2006).

Detection of TEs in the *M. persicae* genome. A 30 000 bp region encompassing each DLS was scanned for adjacent TEs or repetitive sequences using CENSOR (Rebase; <http://www.girinst.org/censor/index.php>) with a reference collection of sequence repeats (Kohany *et al.*, 2006). BLASTX and tBLASTx queries on these flanking DLS regions against the non-redundant (nr) protein database at NCBI were also performed. As a control, repetitive sequences in DLS-free regions of scaffolds harbouring DLSs and on scaffolds where no DLSs could be detected were also analysed.

Transcription of DLSs in *M. persicae*. Transcription of DLSs was analysed by real-time RT-PCR in fourth-instar individuals of *M. persicae* (clone Colmar). Total RNA was extracted from aphids cured of MpDNV (four biological replicates) ground in liquid nitrogen following the Qiagen RNeasy Plant Mini kit protocol (animal tissue protocol). On-column DNase treatment was carried out using an RNase-Free DNase set (Promega). Synthesis of cDNA (from 2 μg RNA) was primed using oligo(dT) (Promega) and cDNA

synthesis was done using the Moloney murine leukemia virus reverse transcriptase (Promega). The cDNA was then used as a template for real-time PCR amplification (primers listed in Table S4) using an iTaq universal SYBR Green supermix (Bio-Rad Laboratories). Ribosomal proteins RLP7 and L27 were used as internal controls. The amplification reaction consisted of $10 \times$ SYBR Green supermix and 10 pmol forward and reverse primers (Table S4) in a final volume of 20 μl . Three technical repeats were performed using a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories). The mRNA synthesis was normalized with RLP7 and L27 using the $2^{-\Delta\Delta\text{CT}}$ method (Livak & Schmittgen, 2001) and analysed with Bio-Rad CFX Manager software (Bio-Rad Laboratories).

ACKNOWLEDGEMENTS

DNA sequence data were downloaded from AphidBase (<http://www.aphidbase.com/aphidbase/>). Funding for *M. persicae* clone G006 genomic sequencing was provided by USDA-NIFA award 2010-65105-20558 held by Alex Wilson (Miami University, OH, USA). Funding for *M. persicae* clone_O genomic sequencing was provided by The Genome Analyses Centre (TGAC) Capacity and Capability Challenge program project CCC-15 held by Saskia Hogenhout and from the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/J004553/1) and the John Innes Foundation. We are grateful to Fabrice Legeai and Antony Bretaudeau (INRA, Rennes) for their assistance in aphid genomic sequences analysis. We also thank Denis Filloux (INRA, Montpellier, France) for his help in database analysis, Neta Luria and Aviv Dombrovsky (Volcani Center, Israel) and Pierre Gladieux (INRA, Montpellier, France) for their fruitful discussions. We are grateful to Keith Garrisson (Saint Mary's College of California, CA, USA) and A. Wilson for critically reading this manuscript.

REFERENCES

- Barreau, C., Jousset, F. X. & Bergoin, M. (1996). Pathogenicity of the *Aedes albopictus* parvovirus (AaPV), a denso-like virus, for *Aedes aegypti* mosquitoes. *J Invertebr Pathol* **68**, 299–309.
- Belyi, V. A., Levine, A. J. & Skalka, A. M. (2010). Unexpected inheritance: multiple integrations of ancient Bornavirus and Ebolavirus/Marburgvirus sequences in vertebrate genomes. *PLoS Pathog* **6**, e1001030.
- Bergoin, M. & Tijssen, P. (2010). Densovirus: a highly diverse group of arthropod Parvoviruses. In *Insect Virology*, pp. 59–82. Edited by S. Asgari & K. Johnson. Norfolk, UK: Academic Press.
- Berns, K. I. (1990). Parvovirus replication. *Microbiol Rev* **54**, 316–329.
- Bertsch, C., Beuve, M., Dolja, V. V., Wirth, M., Pelsy, F., Herrbach, E. & Lemaire, O. (2009). Retention of the virus-derived sequences in the nuclear genome of grapevine as a potential pathway to virus resistance. *Biol Direct* **4**, 21.
- Bézier, A., Annaheim, M., Herbinière, J., Wetterwald, C., Gyapay, G., Bernard-Samain, S., Wincker, P., Roditi, I., Heller, M. & other authors (2009). Polydnviruses of braconid wasps derive from an ancestral nudivirus. *Science* **323**, 926–930.
- Buchatskiĭ, L. P., Kuznetsova, M. A., Lebedinets, N. N. & Kononko, A. G. (1987). [Development and basic properties of the viral preparation viroden]. *Vopr Virusol* **32**, 729–733 (in Russian).
- Carlson, J., Suchman, E. & Buchatsky, L. (2006). Densovirus for control and genetic manipulation of mosquitoes. *Adv Virus Res* **68**, 361–392.
- Cheng, R. L., Xi, Y., Lou, Y. H., Wang, Z., Xu, J. Y., Xu, H. J. & Zhang, C. X. (2014). Brown planthopper nudivirus DNA integrated in its host genome. *J Virol* **88**, 5310–5318.

- Chevenet, F., Brun, C., Bañuls, A. L., Jacq, B. & Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 7, 439.
- Chilana, P., Sharma, A. & Rai, A. (2012). Insect genomic resources: status, availability and future. *Curr Sci* 102, 571–580.
- Croizier, L., Jousset, F. X., Veyrunes, J. C., López-Ferber, M., Bergoin, M. & Croizier, G. (2000). Protein requirements for assembly of virus-like particles of *Junonia coenia* densovirus in insect cells. *J Gen Virol* 81, 1605–1613.
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V. & other authors (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36 (Web Server), W465–W469.
- Ding, C., Urabe, M., Bergoin, M. & Kotin, R. M. (2002). Biochemical characterization of *Junonia coenia* densovirus nonstructural protein NS-1. *J Virol* 76, 338–345.
- Dupont, F. (2003). Risk assessment of the use of autonomous parvovirus-based vectors. *Curr Gene Ther* 3, 567–582.
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792–1797.
- El-Far, M., Li, Y., Fédière, G., Abol-Ela, S. & Tijssen, P. (2004). Lack of infection of vertebrate cells by the densovirus from the maize worm *Mythimna loreyi* (MIDNV). *Virus Res* 99, 17–24.
- Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet* 5, 103–107.
- Flegel, T. W. (2009). Hypothesis for heritable, anti-viral immunity in crustaceans and insects. *Biol Direct* 4, 32.
- Goic, B., Vodovar, N., Mondotte, J. A., Monot, C., Frangeul, L., Blanc, H., Gausson, V., Vera-Otarola, J., Cristofari, G. & Saleh, M. C. (2013). RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model *Drosophila*. *Nat Immunol* 14, 396–403.
- Hu, Y. Y., Zhang, J. M., Yin, Y. N., Wu, K. L. & He, M. (2000). Pathogenicity island: a novel pesticide enhanced by insect hormone for controlling cockroaches. *Chin J Hyg Insect Equipm* 6, 29–32.
- International Aphid Genomics Consortium (2010). Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol* 8, e1000313.
- Jiang, H., Zhang, J. M., Wang, J. P., Yang, B., Liu, C. F., Lu, J. & Hu, Y. Y. (2007). Genetic engineering of *Periplaneta fuliginosa* densovirus as an improved biopesticide. *Arch Virol* 152, 383–394.
- Jones, D. T., Taylor, W. R. & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8, 275–282.
- Katzourakis, A. & Gifford, R. J. (2010). Endogenous viral elements in animal genomes. *PLoS Genet* 6, e1001191.
- Kohany, O., Gentles, A. J., Hankus, L. & Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 7, 474.
- Krupovic, M. & Forterre, P. (2015). Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. *Ann N Y Acad Sci* 1341, 41–53.
- Legendre, D. & Rommelaere, J. (1992). Terminal regions of the NS-1 protein of the parvovirus minute virus of mice are involved in cytotoxicity and promoter *trans* inhibition. *J Virol* 66, 5705–5713.
- Liu, H., Fu, Y., Li, B., Yu, X., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Yi, X. & Jiang, D. (2011a). Widespread horizontal gene transfer from circular single-stranded DNA viruses to eukaryotic genomes. *BMC Evol Biol* 11, 276.
- Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Peng, Y., Yi, X. & Jiang, D. (2011b). Widespread endogenization of densoviruses and parvoviruses in animal and human genomes. *J Virol* 85, 9863–9876.
- Liu, K., Li, Y., Jousset, F. X., Zadori, Z., Szelei, J., Yu, Q., Pham, H. T., Lépine, F., Bergoin, M. & Tijssen, P. (2011c). The *Acheta domesticus* densovirus, isolated from the European house cricket, has evolved an expression strategy unique among parvoviruses. *J Virol* 85, 10069–10078.
- Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25, 402–408.
- Maori, E., Tanne, E. & Sela, I. (2007). Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. *Virology* 362, 342–349.
- Monde, K., Contreras-Galindo, R., Kaplan, M. H., Markovitz, D. M. & Ono, A. (2012). Human endogenous retrovirus K Gag coassembles with HIV-1 Gag and reduces the release efficiency and infectivity of HIV-1. *J Virol* 86, 11194–11208.
- Monsarrat, P., Mariau, D. & Genty, P. (1984). Densovirus en lutte biologique. *Bull Soc Entomol Fr* 89, 816–821.
- Mutuel, D., Ravallec, M., Chabi, B., Multeau, C., Salmon, J. M., Fournier, P. & Ogliastro, M. (2010). Pathogenesis of *Junonia coenia* densovirus in *Spodoptera frugiperda*: a route of infection that leads to hypoxia. *Virology* 403, 137–144.
- Ramsey, J. S., Wilson, A. C., de Vos, M., Sun, Q., Tamborindeguy, C., Winfield, A., Malloch, G., Smith, D. M., Fenton, B. & other authors (2007). Genomic resources for *Myzus persicae*: EST sequencing, SNP identification, and microarray design. *BMC Genomics* 8, 423.
- Rhode, S. L., III & Richard, S. M. (1987). Characterization of the *trans*-activation-responsive element of the parvovirus H-1 P38 promoter. *J Virol* 61, 2807–2815.
- Rozen, S. & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132, 365–386.
- Ryabov, E. V., Keane, G., Naish, N., Evered, C. & Winstanley, D. (2009). Densovirus induces winged morphs in asexual clones of the rosy apple aphid, *Dysaphis plantaginea*. *Proc Natl Acad Sci U S A* 106, 8465–8470.
- Suto, C. (1979). Characterization of a virus newly isolated from the smoky-brown cockroach, *Periplaneta fuliginosa* (Serville). *Nagoya J Med Sci* 42, 13–25.
- Szelei, J., Woodring, J., Goettel, M. S., Duke, G., Jousset, F. X., Liu, K. Y., Zadori, Z., Li, Y., Styer, E. & other authors (2011). Susceptibility of North-American and European crickets to *Acheta domesticus* densovirus (AdDNV) and associated epizootics. *J Invertebr Pathol* 106, 394–399.
- Tal, J. & Attathom, T. (1993). Insecticidal potential of insect parvovirus GmDNV. *Arch Insect Biochem Physiol* 22, 345–356.
- Thézé, J., Bézier, A., Periquet, G., Drezen, J. M. & Herniou, E. A. (2011). Paleozoic origin of insect large dsDNA viruses. *Proc Natl Acad Sci U S A* 108, 15931–15935.
- van Munster, M., Dullemans, A. M., Verbeek, M., van den Heuvel, J. F., Reinbold, C., Brault, V., Clériver, A. & van der Wilk, F. (2003a). Characterization of a new densovirus infecting the green peach aphid *Myzus persicae*. *J Invertebr Pathol* 84, 6–14.
- van Munster, M., Dullemans, A. M., Verbeek, M., van den Heuvel, J. F., Reinbold, C., Brault, V., Clériver, A. & van der Wilk, F. (2003b). A new virus infecting *Myzus persicae* has a genome organization similar to the species of the genus *Densovirus*. *J Gen Virol* 84, 165–172.
- Xu, P., Liu, Y., Graham, R. I., Wilson, K. & Wu, K. (2014). Densovirus is a mutualistic symbiont of a global crop pest (*Helicoverpa armigera*) and protects against a baculovirus and Bt biopesticide. *PLoS Pathog* 10, e1004490.
- Yang, C. C., Xiao, X., Zhu, X., Ansardi, D. C., Epstein, N. D., Frey, M. R., Matera, A. G. & Samulski, R. J. (1997). Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration in vivo and in vitro. *J Virol* 71, 9231–9247.