

# Evolution and phylogeography of *Culex pipiens* densovirus

Mine Altinli,<sup>1,\*†‡</sup> Sebastian Lequime,<sup>2</sup> Maxime Courcelle,<sup>1</sup> Sarah François,<sup>3,4</sup> Fabienne Justy,<sup>1</sup> Anne-Sophie Gosselin-Grenet,<sup>3</sup> Mylene Ogliastro,<sup>3</sup> Mylene Weill,<sup>1</sup> and Mathieu Sicard<sup>1,\*</sup>

<sup>1</sup>ISEM, Université de Montpellier, CNRS, IRD, EPHE, Montpellier, France, <sup>2</sup>KU Leuven, Department of Microbiology, Immunology, and Transplantation, Laboratory of Clinical and Epidemiological Virology, Rega Institute, Leuven, Belgium, <sup>3</sup>DGIMI, INRA, Université de Montpellier, Montpellier, France and <sup>4</sup>Department of Zoology, University of Oxford, Oxford, UK

\*Corresponding authors: E-mails: altinlimine@gmail.com and mathieu.sicard@umontpellier.fr

†Present address: Molecular Entomology Team, Molecular Biology and Immunology Department, Bernhard Nocht Institute for Tropical Medicine, Hamburg 20359, Germany.

‡<https://orcid.org/0000-0001-8557-9442>

## Abstract

Viruses of the *Parvoviridae* family infect a wide range of animals including vertebrates and invertebrates. So far, our understanding of parvovirus diversity is biased towards medically or economically important viruses mainly infecting vertebrate hosts, while invertebrate infecting parvoviruses—namely densoviruses—have been largely neglected. Here, we investigated the prevalence and the evolution of the only mosquito-infecting ambidensovirus, *Culex pipiens* densovirus (CpDV), from laboratory mosquito lines and natural populations collected worldwide. CpDV diversity generally grouped in two clades, here named CpDV-1 and -2. The incongruence of the different gene trees for some samples suggested the possibility of recombination events between strains from different clades. We further investigated the role of selection on the evolution of CpDV genome and detected many individual sites under purifying selection both in non-structural and structural genes. However, some sites in structural genes were under diversifying selection, especially during the divergence of CpDV-1 and -2 clades. These substitutions between CpDV-1 and -2 clades were mostly located in the capsid protein encoding region and might cause changes in host specificity or pathogenicity of CpDV strains from the two clades. However, additional functional and experimental studies are necessary to fully understand the protein conformations and the resulting phenotype of these substitutions between clades of CpDV.

**Key words:** densovirus evolution; *Culex pipiens*; insect-specific virus prevalence; molecular evolution; phylogeography; selection.

## 1. Introduction

Single-stranded DNA (ssDNA) viruses are widespread and infect hosts from all domains of life and environmental conditions (Pietilä et al. 2009; Rosario and Breitbart 2011; Mochizuki et al.

2012; Krupovic 2013). Their integration in many of their hosts' genomes as endogenous viral elements also suggests long-term evolution with their hosts (Belyi, Levine, and Skalka 2010; Kapoor, Simmonds, and Lipkin 2010; Liu et al. 2011; Thézé et al.

2014; Metegnier et al. 2015). One of the most widely distributed ssDNA virus family of animals, the *Parvoviridae* family can infect both vertebrate (*Parvovirinae* subfamily) and invertebrate hosts (*Densovirinae* subfamily) (Cotmore et al. 2014, 2019). Furthermore, the recent advances in high-throughput sequencing technologies have unveiled an unexpected presence and diversity of parvoviruses, particularly in *Densovirinae* subfamily (i.e. densovirus) (François et al. 2016). This subfamily so far consists of five genera (i.e. *Iteradensovirus*, *Brevidensovirus*, *Ambidensovirus*, *Penstylidensovirus*, and *Hepandensovirus*) isolated mostly from insects (e.g. lepidopteran, dipteran, orthopteran, and dictyopteran), crustaceans (e.g. prawns and crayfish) and more recently from echinoderms (e.g. sea urchins and sea stars) (Cotmore and Tattersall 2014; Gudenauf et al. 2014; Hewson et al. 2014; Tijssen et al. 2016; Cotmore et al. 2019).

All parvoviruses have a small non-enveloped capsid (18–26 nm), protecting a 4–6 kb ssDNA linear genome which encodes for two sets of proteins: non-structural proteins (NS) involved in the regulation of gene expression and genome replication, and structural proteins (VP), the components of the viral capsids (Cotmore and Tattersall 2014; Cotmore et al. 2019). The *Parvoviridae* family, like most ssDNA viruses, is thought to originate from ancient recombination events resulting in the acquisition of these ns genes from DNA donors (such as bacterial plasmids) and vp genes from RNA viruses (Krupovic 2013; Koonin, Dolja and Krupovic 2015). Recombination also plays an essential part in the evolution among parvoviruses, with breakpoints generally positioned between the ns and the vp genes that in many cases seemed to result in chimaeras with ns and vp genes of different origins (Shackelton et al. 2007; Lefeuvre et al. 2009; Martin et al. 2011; Xu et al. 2012; Leal et al. 2013).

In addition to frequent recombination, parvovirus genome evolution seems to be characterised by high substitution rates (Shackelton et al. 2005, 2006; Firth et al. 2009; Truyen et al. 2011; Stamenković et al. 2016). The variations in substitution rates between viruses are usually explained by whether they use high-fidelity DNA polymerase or low-fidelity RNA polymerases (Elena and Sanjuán 2007; Duffy, Shackelton, and Holmes 2008). Although parvovirus replication relies on host DNA polymerases, their substitution rates  $10^{-3}$  to  $10^{-5}$  substitutions/site/year are closer to similarly sized RNA viruses ( $10^{-2}$  to  $10^{-5}$  substitutions/site/year) than to large double-stranded DNA viruses ( $10^{-5}$  to  $10^{-9}$  substitutions/site/year; Shackelton et al. 2005, 2006; Duffy, Shackelton, and Holmes 2008; Firth et al. 2009; Truyen et al. 2011; Stamenković et al. 2016). Other factors can also affect the substitution rate variations, such as the generation time, transmission, and selection modes (Duffy, Shackelton, and Holmes 2008). Although a general purifying selection is observed in parvoviruses (Truyen et al. 2011; Martynova, Schal, and Mukha 2016; Stamenković et al. 2016), diversifying selection combined with fast mutation rates can result in the expansion of the host range and the emergence of new viral species (Shackelton et al. 2005). For instance, the emergence of the canine parvovirus from the feline panleukopenia parvovirus, when the virus jumped from cats to dogs, was associated with a very high substitution rate ( $10^{-3}$ ) and high positive selection pressure (Shackelton et al. 2005).

In contrast to parvoviruses, our knowledge of densovirus genome evolution is very limited. Only one study highlighted the impact of recombination and purifying selection on densovirus evolutionary history (Martynova, Schal, and Mukha 2016). A better understanding of the evolution of densovirus thus requires more studies of their diversity in nature. Among densovirus, those infecting mosquitoes (MDVs) have been

isolated from several species in both natural populations and laboratory colonies or from mosquito-derived cell lines (Afanasiev, Galyov, and Buchatsky 1991; Jousset et al. 1993; Boublik, Jousset, and Bergoin 1994; O'Neill et al. 1995; Kittayapong, Baisley, and O'Neill 1999; Jousset, Baquerizo, and Bergoin 2000; Rwegoshora, Baisley, and Kittayapong 2000; Ng et al. 2011; Boonak et al. 2015). Most of the MDVs discovered so far are closely related (82–99% sequence homology) and considered to be different strains from two distinct species (i.e. Dipteran *Brevidensovirus* 1 and 2) that belong to the *Brevidensovirus* genus (Cotmore et al. 2014, 2019). The *Culex pipiens* densovirus (CpDV) isolated from *C. pipiens* mosquitoes is the only known exception, as it belongs to the *Ambidensovirus* genus due to its ambisense genomic structure (Cotmore et al. 2014, 2019). CpDV has a large genome (i.e. ~6 vs ~4 kb for brevidensoviruses) putatively encoding for three NS and four VP proteins (Jousset, Baquerizo and Bergoin 2000; Baquerizo-Audiot et al. 2009; Cotmore et al. 2014, 2019).

First isolated two decades ago following an unusually high larval mortality in *C. pipiens* colonies, CpDV still persists in the same laboratory colonies in all life stages of mosquitoes without causing high mortality (Jousset, Baquerizo, and Bergoin 2000; Altinli et al. 2019). This covert infection in rearing colonies might result from facilitation of horizontal transmission in laboratory conditions, in addition to the observed vertical transmission (Altinli et al. 2019). However, whether and how this virus persists and evolves in natural populations were unknown. To investigate these questions, we designed a CpDV-specific PCR-based diagnostic test and sequencing assays to analyse CpDV prevalence in nature, and their diversity in *C. pipiens* mosquitoes collected from both natural populations and laboratory colonies over 30 years.

## 2. Methods

### 2.1 Biological Material

Samples from natural *C. pipiens* (s.l.) populations have been collected between 1975 and 2017 (Guillemaud, Pasteur, and Rousset 1997; Pasteur, Nancé, and Bons 2001; Duron et al. 2005; Dumas et al. 2013; Pocquet et al. 2013; Atyame et al. 2015; Altinli, Gunay, and Alten 2018; Supplementary Table S1). DNA has been extracted from individual mosquitoes by cetyltrimethylammonium bromide (CTAB; Rogers and Bendich 1989), or phenol/chloroform methods (Green and Sambrook 1989), as previously described in Dumas et al. (2013), Atyame et al. (2015), Pocquet et al. (2013), and Altinli et al. (2018) and kept at  $-80^{\circ}\text{C}$  since then. Additionally, larvae from natural populations were collected in Armenia, Benin, France, Italy, and Ivory Coast (Supplementary Table S1) and their DNA extracted using CTAB method.

To follow the genetic variations between CpDV strains in the laboratory colonies, we extracted DNA from adult mosquitoes from one of our colonies, Slab. The Slab colony has been initially established with mosquitoes sampled in the USA in the 1960s (Georghiou, Metcalf and Gidden 1966). It has been reared in our insectary (ISEM, Montpellier, France) for at least 30 years until present, and regularly sampled and stored in liquid nitrogen from 1990 to 2018. DNA was extracted using Qiagen extraction kit according to the manufacturer's instructions. CpDV reference genome has been sequenced from a sample isolated from the same Slab line in 1994, hereinafter mentioned as Reference\_1994 (Jousset, Baquerizo, and Bergoin 2000). Isolated CpDV suspension has been stored in  $-80^{\circ}\text{C}$

since then and used as a positive control for PCR assays in this study.

## 2.2 CpDV presence in natural populations

DNA from 2,843 individual mosquitoes collected from natural populations from 136 different locations were tested for CpDV presence by PCR (Supplementary Table S1). These samples did not include the samples from laboratory colonies. A CpDV-specific set of primers was designed to amplify 238 bp (from 1606 to 1843, Supplementary Table S2 for primers and PCR protocols) where NS2 and NS2' open reading frames (ORFs) overlap, as such overlapping regions are expected to be highly conserved (Fig. 1).

## 2.3 Partial sequencing of CpDV genomes

To investigate the diversity of CpDV, we have designed primers both in the NS1 and VP coding regions using CpDV sequences obtained from a pilot viral metagenomic study. For that, *C. pipiens* larvae pools from four populations (Montpellier and Benin natural populations, Utique and Slab laboratory mosquito lines) were processed and sequenced as previously described (François et al. 2018). Taxonomic assignment was obtained through searches against the NCBI RefSeq viral database with an e-value cut-off of  $<10^{-3}$ . CpDV contigs quality was assessed by mapping using Bowtie 2.1.0 (options end-to-end very sensitive). Mapped reads, along with the reference genome (Baquerizo-Audiot et al. 2009) were used to create a consensus genome using SeaView v4.7 (Gascuel, Gouy, and Lyon 2010). We used the resultant consensus genome to design CpDV-specific primers of which the specificity was verified using primer-BLAST (Ye et al. 2012).

Semi-nested PCR protocols were designed with CpDV-specific primers to amplify 967 bp within NS1 coding region (1681–2648; Fig. 1) and a 1,389 bp within VP coding region (3470–4959; Fig. 1; Supplementary Table S2 for primers and PCR protocols). NS and VP amplicons together covered about 40% of the CpDV genome (2,356 over 6000 bp) and encompassed the conserved SF3 helicase domain classically used for phylogenetic construction of parvoviruses as well as the conserved PLA2 domain (Fig. 1; Cotmore et al. 2014; Mietzsch, Pénczes, and Agbandje-Mckenna 2019). For sequencing, we chose the samples that showed a strong band on agarose gel with the diagnostic test, and aimed to cover the diversity among different continents as

much as possible. We also sequenced Slab samples to follow the evolution of CpDV in the laboratory colony. All the sequences are publicly available on GenBank, and their respective accession numbers are given in Supplementary Table S3.

## 2.4 Sequence similarity

We compared NS1 and VP protein partial sequences from all the samples (see Supplementary Material for the alignments) and calculated their identity scores using online tool SIAS (sequences Identities and Similarities, <http://imed.med.ucm.es/Tools/sias.html>). A normalised similarity score was calculated based on BLOSUM62 substitution matrix with default gap penalty.

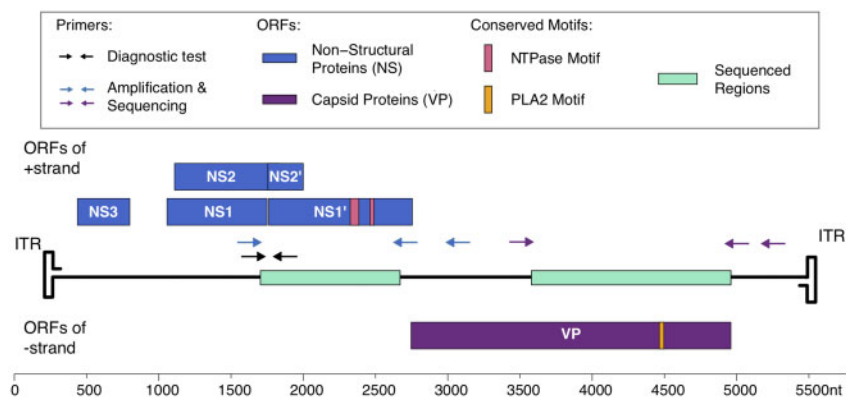
## 2.5 Recombination

We checked for recombination in both *ns1* and *vp* datasets separately as well as the concatenated *ns1-vp* dataset, using GARD (Genetic Algorithm for Recombination Detection) package on DATAMONKEY web server (Pond et al. 2006; Weaver et al. 2018).

## 2.6 Selection

*ns1* and *vp* ORFs were analysed for selection using HyPhy package (Pond, Frost, and Muse 2005) on DATAMONKEY web server. Selection was inferred by the ratio ( $\omega$ ) of non-synonymous substitutions (dN) over synonymous substitutions (dS) for a given site, where  $\omega > 1$ ,  $\omega = 1$ ,  $\omega < 1$  represent diversifying (positive), neutral and purifying (negative) selection, respectively.

We used MEME (mixed effects model of evolution) (Murrell et al. 2012) and FEL (fixed effects likelihood) (Kosakovsky Pond and Frost 2005) to test for episodic (occurring in only some of the branches of the phylogeny) and pervasive selection (occurring in the whole tested phylogeny) acting on individual sites, respectively. MEME approach detects sites under episodic diversifying selection ( $dN/dS > 1$ ). MEME infers a single  $\alpha$  (dS) and two separate  $\beta$  (dN) values for each site. One of the  $\beta$  values ( $\beta^+$ ) is constrained in the null model and unconstrained in the alternative model. A given site is considered under diversifying selection if its  $\beta^+ > \alpha$  is significant using the likelihood ratio test. FEL approach fits an MG94xREV model to each site, after the optimisation of branch lengths and nucleotide substitution parameters, and infers dN and dS rates on each site. Likelihood ratio test (LRT) was then used to test if a given dN is significantly higher



**Figure 1.** CpDV Genome encodes for two types of proteins: non-structural (NS) and structural (VP). CpDV genome is ambisense with ORFs of NS and VP proteins situated on complementary strands. CpDV has three putative NS proteins as previously annotated (Baquerizo-Audiot et al. 2009): NS3, NS2, and NS1 (ORFs are shown as blue boxes). CpDV's VP proteins are putatively encoded by a single ORF (shown in purple) and leaky scanning. We have sequenced partial *ns1* (blue arrows) and *vp* (purple arrows) genes to infer CpDV diversity, covering the conserved SF3 helicase domain of NS1 consisting of Walker A and B boxes (NTPase motif, pink boxes), and PLA2 of VP (orange box). Location of primers either used in semi-nested PCRs or for the diagnostic test has been shown as arrows.

than the dS and a *P*-value is derived. If the *P*-value was significant ( $P < 0.05$ ), the site was classified based on whether  $dN > dS$  (diversifying selection) or  $dN < dS$  (purifying selection).

To test whether selection played a role in the diversification of major clades, we used a branch-site model, aBSREL (adaptive Branch-Site Random Effects Likelihood) approach (Smith et al. 2015). aBSREL assigns an optimal number of  $\omega$  rate classes for each branch using Akaike Information Criterion<sub>c</sub> (small sample AIC) and fits the full adaptive model. Then, for each branch, the full model was compared with the null model where branches were not allowed to have rate classes of  $\omega > 1$  with the LRT. We defined the different branches between major clades (e.g. CpDV-1 and -2) to investigate whether a significant proportion of sites have evolved under diversifying selection. As aBSREL does not test for selection at specific sites, we also used FEL approach on the same branches to study the selection on individual sites.

To visualise the individual sites under positive selection, their location in VP protein and their possible effect on the protein structure, we used the structural information from the crystal structure of the VP from *Galleria melonella* densovirus (GmDV) (Simpson et al. 1998). Consensus sequences from CpDV-1 to -2 clades were aligned, along with some representatives from these clades to highlight the amino acid changes both between and within CpDV clades, as well as the changes between CpDV and VP4 protein of GmDV. This alignment was then visualised and combined with the secondary structure information from PDB using ESPript 3.0 (Robert and Gouet 2014).

## 2.7 Phylogenetics and phylogeography of CpDV

The presence of a clock signal was assessed using TempEst v1.5 with optimised rooting (Rambaut et al. 2016). A positive correlation between sampling time and root-to-tip divergences was obtained although correlation coefficients were low for both NS1 and VP. To ensure the clock signal, we randomly swapped the collection dates between samples and estimated the rate of evolution ('null' distribution of rate estimates) as previously described (Trovão et al. 2015). A clear overlap between the credible interval of mean evolutionary rates (calculated as 95% highest posterior density (HPD)) of the data with random and the original collection dates proved the lack of a true clock signal in the data (Supplementary Table S4). Thus, we were not able to estimate time to the most recent common ancestor (tMRCA) and evolutionary rates for CpDV strains.

As the data did not have enough clock signal, we ran both a constant population size model and an exponential growth model, by fixing the evolutionary rate in BEAST v1.10.4 (Suchard et al. 2018). Bayesian phylogenetic trees were constructed with HKY85+G substitution model, the best fitting substitution model using Modelfinder (Kalyaanamoorthy et al. 2017). As 95% HPD of the growth rate in the exponential growth model included 0, it was not statistically different from a constant population size model (Supplementary Material).

To assess the phylogeographic spread of CpDV in space, we performed a discrete phylogeographic reconstruction. We grouped the samples from the same geographic regions to decrease the sampling bias (e.g. Turkey and Armenia as Near-East; China, Vietnam, Thailand as Eastern Asia; Sub-Saharan African countries as Sub-Saharan Africa). As the origin of the CpDVs in our laboratory lines was unclear we ran the models excluding these samples (i.e. Slab samples). Discrete location transition of CpDV between the identified ten geographical regions was then modelled using a reversible symmetric continuous-time Markov chain (CTMC) process (Lemey et al. 2009). Each CTMC analysis was set in three

independent replicates to ensure proper convergence and ran for 50 million iterations sampled every 5,000 generations. Convergence (Effective Sample Size  $> 200$ ) and mixing of all parameters were verified using Tracer v1.7 (Rambaut et al. 2018). The first 10% of the samples were removed as burn-in, and the log and tree files were merged using LogCombiner v1.10.4 (Suchard et al. 2018).

The phylogeographic analysis was performed using TreeAnnotator v1.10.4 (Suchard et al. 2018) to allow the summary of location estimates on a Maximum Clade Credibility (MCC) tree after a 10% burn-in was discarded, and then visualised in R using the packages ape (Paradis and Schliep 2018) and ggtree (Yu et al. 2017), a wrapper for ggplot2 (Wickham 2009) designed to handle phylogenetic trees.

## 3. Results

### 3.1 CpDV prevalence in natural populations of *C. pipiens* worldwide

To investigate the general prevalence of CpDV in natural *C. pipiens* (s.l.) populations, DNA samples from a total of 2,843 individual mosquitoes collected between 1975 and 2017 from 136 different locations were subjected to a PCR-based CpDV-specific diagnostic test (Fig. 1 and Supplementary Table S1). 48% of the samples were CpDV-positive (Fig. 2 and Supplementary Table S1), displaying an amplicon length at the expected size. We did not detect CpDV in any of the samples from 29 localities out of 136.

### 3.2 Worldwide diversity and evolution of CpDV

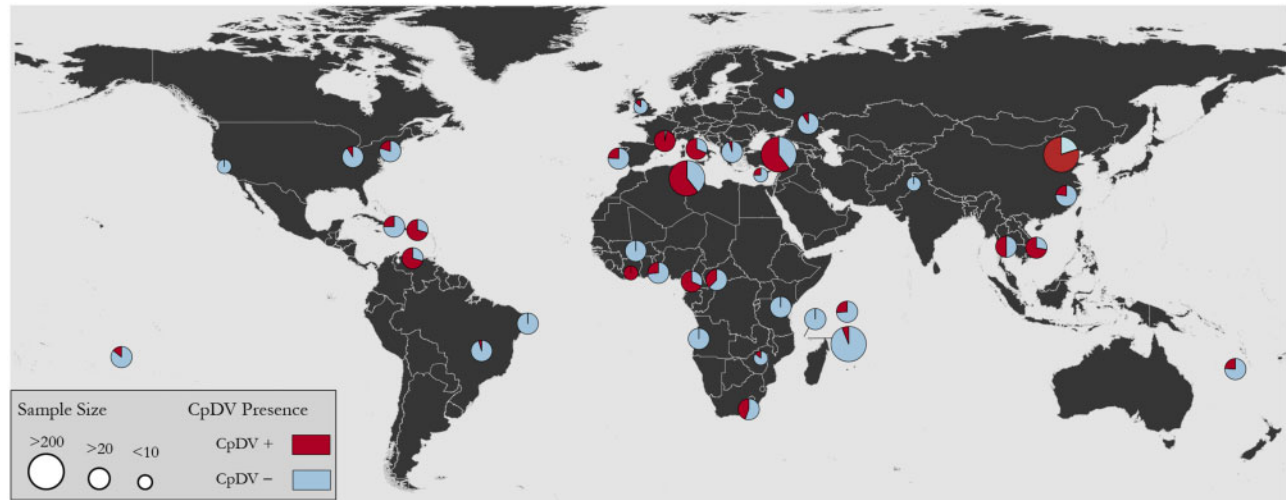
To analyse the diversity and evolution of CpDV, a total of fifty-one samples (i.e. individual mosquitoes) were sequenced for *ns1* ORF. Thirty-six samples out of these fifty-one were successfully amplified and sequenced for *vp* ORF (Supplementary Table S3). We could not obtain *vp* amplicons for fifteen samples probably because of 1) a possibly higher diversity in this region, preventing primers to be suitable for all the samples, or 2) poor DNA conservation over the years and/or the low CpDV titres.

CpDV strains were highly similar both for NS1 (mean 98%, min = 94%, max = 100%) and VP sequences (mean = 96%, min = 83%, max = 100%). This confirmed that all the sampled strains belong to *Dipteran Ambidensovirus 1* species as densoviruses are taxonomically assigned at species level based on the identity criterium of the NS1 protein sequence with a  $< 85\%$  cut off (Cotmore et al. 2019).

#### 3.2.1 Recombination

We checked for recombination events in CpDV considering both *ns1* and *vp* datasets separately and a concatenation of *ns1-vp* datasets. We did not detect any recombination within *ns1* dataset. Although a recombination event was detected in the *vp* gene (GARD,  $P < 0.001$ ), this did not affect the phylogenetic reconstruction significantly. Contrarily, recombination events detected in the concatenated dataset (GARD,  $P < 0.001$ ) affected phylogenetic tree significantly. Therefore, *ns1* and *vp* phylogenetic trees were reconstructed separately. To investigate this possible breakpoint between *ns1* and *vp*, we further designed new primers within our *ns1-vp* sequences. However, we were able to amplify this region only for fresh samples but not for putative recombinant samples, probably due to the combination of poor DNA quality and the size of the sequence (~2,000 bp).





**Figure 2.** Worldwide prevalence of CpDV in natural *C. pipiens* (s.l.) populations A total of 2,843 individual larvae collected from 136 different locations worldwide were tested with a CpDV PCR-based diagnostic test designed in a conserved region of the *ns2* gene. As sampling sites were too numerous for some specific areas (Turkey, Tunisia, and Mayotte) we grouped them for clarity in this figure (see [Supplementary Table S1](#) for the full list of sampling locations, years, and CpDV prevalence). In total 48% of these samples were positive, representing 107 locations out of 136 locations tested that is, no amplification was obtained for individuals collected from 29 locations.

### 3.2.2 Phylogeny of CpDV

The tree corresponding to *ns1* gene, showed three distinct and well-supported clades, tentatively named CpDV-1, -2, and -3 (Fig. 3A). CpDV-3 clade was represented only by two samples collected from Beijing in 2003 (Fig. 3A). The other two clades, CpDV-1 and -2, correspond to samples collected at different times and geographical locations (Fig. 3A).

*ns1* and *vp* gene phylogenetic trees were similar in topology, although the *vp* gene tree discriminated less clades (only CpDV-1 and -2; Fig. 3B). This lack of CpDV-3 clade in the *vp* tree is most likely due to the lack of amplification of the China-1 sample (Fig. 3A), and/or to the fact that VP region of China-2 belonged to CpDV-1 clade because of a putative recombination event (indicated in red Fig. 3A). Indeed, as suggested by the detection of recombination events in the concatenated dataset, three samples belonged to different clades depending on the gene used to build the tree (i.e. Slab 1997, China-2 and Puerto Rico-1; Fig. 3, shown in red).

### 3.2.3 CpDV evolution in laboratory colonies

We followed the evolution of CpDV strains sampled over time (including the Reference\_1994) from the Slab laboratory mosquito colony maintained in Montpellier between 1990 and 2018. Collected within a couple of years, Slab\_1990 and Reference\_1994 were closely related (100% NS1, 97.54% VP) and belonged to the CpDV-2 clade (Fig. 3). Interestingly, after 1997 all samples sequenced from the Slab line belonged to CpDV-1 clade in both gene phylogenetic trees and only differed by few substitutions, except the uncertainty for the *vp* of Slab\_2007, which we could not amplify. The incongruence of *ns1* (CpDV-1) and *vp1* (CpDV-2) genes of Slab\_1997 strain (Fig. 3) suggested that 1) CpDV strains belonging to the two clades co-existed in mosquito individuals from the Slab laboratory line maintained in Montpellier, 2) a recombination event probably occurred in 1997 between viruses from both clades, and 3) viruses from CpDV-1 have replaced those from CpDV-2 over the years.

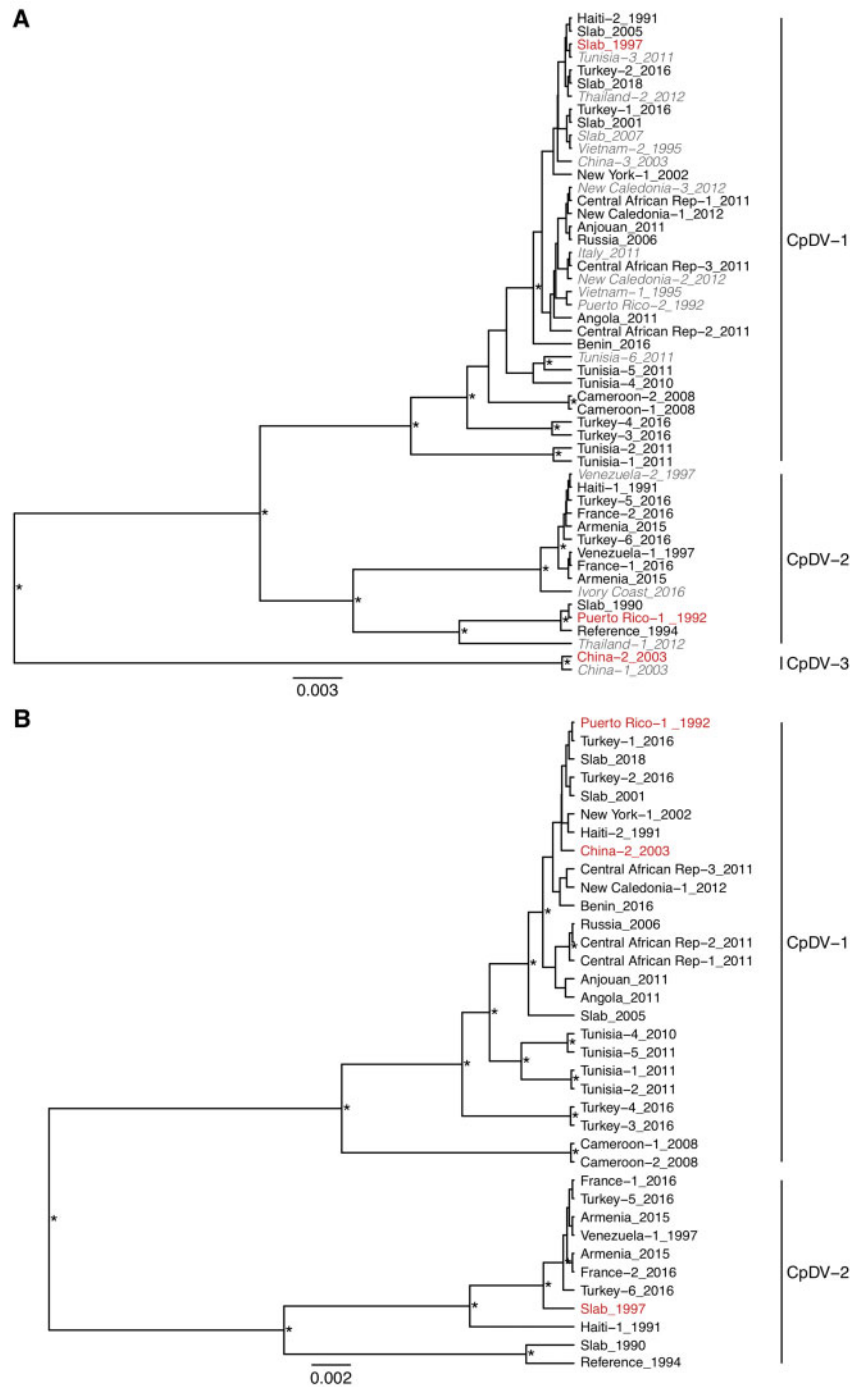
### 3.2.4 Selection

Selection analyses suggested that most of the sites of the CpDV genome are subjected to neutral or purifying selection (Table 1).

Indeed, we detected many sites evolving under pervasive purifying selection in both ORFs: 25 out of 260 sites within NS1 and 21 out of 443 within VP (Table 1, detected with FEL,  $dN/dS < 1$  with  $P < 0.05$ ).

We did not detect any significant diversifying selection for NS1. In VP, two sites were subjected to episodic (i.e. only occurring in some of the branches of the phylogeny) diversifying selection (Fig. 4, sites 343 and 412 in the partial sequences obtained here). Interestingly, diversifying selection detected on Site 343 was caused by one non-synonymous nucleotide substitution between Slab\_1990 and Reference\_1994, resulting in a Glycine to Threonine change (MEME,  $P = 0.009$ ; Fig. 4). Similarly, Site 412 was an Alanine for all strains belonging to CpDV-2 clade and Cameroon 1 and 2 samples but was an Aspartic acid in all the other samples (MEME,  $P = 0.046$ ; Fig. 4).

We further investigated a possible role of positive selection on the diversification of the major clades (i.e. for VP: CpDV-1 and -2, for NS1: CpDV-1 and -2, CpDV-1/2 and -3). To test whether positive selection occurred on a proportion of branches, we used aBSREL which assigns optimal  $dN/dS$  ( $\omega$ ) classes for each branch and compares with a null (constrained) model where branches are not allowed to have rate classes of  $\omega > 1$ . On both trees, none of the branches separating the major clades showed significant differences between optimal and constrained models, suggesting that these branches were not evolving under diversifying selection. For NS1,  $\omega$  distribution was assigned to only one  $\omega$  class both for the branch between CpDV-1/2 and -3 ( $\omega = 0.09$ , for 100% of the sites) and the branch between CpDV-1 and -2 ( $\omega = 0$ , for 100% of the sites). For VP, we observed a small proportion of sites evolving under diversifying selection for CpDV-1/-2 branch ( $\omega_2 = 5.55$ , 8% of sites), even though this proportion was not significant compared with the proportion of the sites under purifying selection ( $\omega_1 = 0$ , 92% of sites,  $P = 0.1$ ). Using FEL approach, we detected six different sites that were likely subjected to diversifying selection (Fig. 4 and Table 2). For these six sites, a non-synonymous change occurred during the diversification of CpDV-1 from CpDV-2 clade, and was conserved in the strains within the clades. Most of the sites (five out of six at Positions 243, 290, 311, 324, and 439) that evolved under diversifying selection were located within a conserved region of VP



**Figure 3.** Bayesian MCC tree of CpDV using (A) fifty-one sequences from the *ns1* ORF and (B) thirty-six sequences from the *vp* ORF. Nodes with high posterior probability support (>0.85) are indicated with an asterisk. The scale indicates nucleotide substitutions per site. The tree was rooted under the assumption of a strict, empirically fixed, molecular clock (see also [Supplementary Material](#) for the annotated tree files). Putative recombinant strains are indicated in red. Samples that were only amplified in *ns1* and not in *vp* are shown in italics and grey.

corresponding to a conserved structural domain (277–375 of GmDV VP4), which is also shared with many vertebrate parvoviruses ([Fig. 4](#)). We also found that the PLA2 and  $\text{Ca}^{2+}$  binding motifs in the *vp* ORF were highly conserved with only one synonymous change in Slab 1990 for the PLA2 motif.

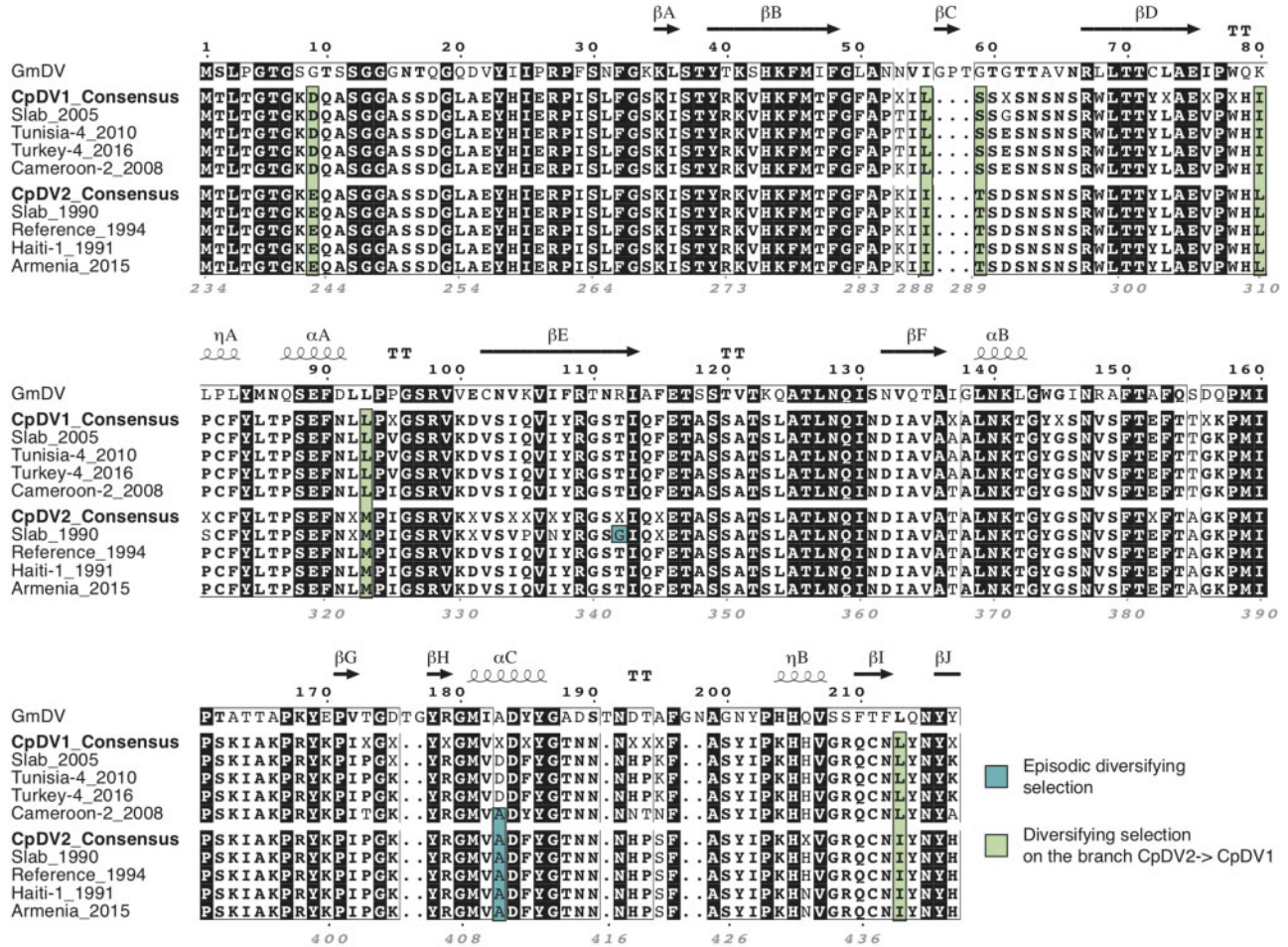
Additionally, the SF3 helicase domain of NS1 including the Walker A and Walker B motifs was also fully conserved, as were the putative start and stop codons of other predicted ORFs within the NS1' sequence.

### 3.3 Phylogeography of CpDV

To estimate the evolutionary rates and the tMRCA of CpDV strains, we first checked the data for a clock signal in using TempEst v1.5 with optimised rooting ([Rambaut et al. 2016](#)). Although we detected a positive correlation between sampling time and root-to-tip divergence, correlation coefficients were low for both *ns1* and *vp* (28 years date range for both NS1 and VP: NS1, 51 sequences, correlation coefficient: 0.29,  $R^2 = 0.08$ ; VP, 29

**Table 1.** Purifying selection acting on individual sites of NS1 and VP: In total, 25 out of 260 sites within NS1 and 21 out of 443 within VP were found to be under purifying selection with  $P < 0.05$ , using FEL. Numbers indicate the position of a given site within the translated protein alignment, the position of the first nucleotide of a given amino acid in the reference CpDV genome is given in brackets.

	Mean rate of dN/dS	Purifying selection on individual sites
NS1	0.137	79 (2008), 81 (2014), 85 (2026), 86 (2029), 98 (2065), 106 (2089), 118 (2125), 123 (2140), 141 (2194), 145 (2206), 159 (2248), 160 (2251), 162 (2257), 164 (2263), 170 (2281), 185 (2326), 189 (2338), 197 (2362), 218 (2425), 224 (2443), 227 (2452), 232 (2467), 247 (2512), 252 (2527), 258 (2545)
VP	0.175	19 (4814), 23 (4802), 24 (4799), 36 (4763), 37 (4760), 38 (4757), 90 (4601), 92 (4595), 93 (4592), 103 (4562), 134 (4469), 141 (4448), 142 (4445), 179 (4334), 265 (4076), 335 (3866), 352 (3815), 371 (3758), 373 (3752), 383 (3722), 396 (3683)



**Figure 4.** Diversifying selection on a conserved protein domain of CpDV capsid protein. Alignment of the domain of VP from CpDV that corresponds to the VP4 from GmDV including the structural domain 'BIDG' sheet ( $\beta B$ ,  $\beta I$ ,  $\beta D$ , and  $\beta G$  strands shown as arrows). As the structure of CpDV capsid is still unknown, the CpDV VP consensus sequences were aligned with the GmDV sequences, of which the capsid structure is resolved. X within a consensus sequence indicates the substitutions within the clade. Some representative sequences from the two clades were also added to the alignment to highlight the amino acid changes within each clade. Sites that are under episodic diversifying selection are highlighted in blue. Non-synonymous substitution between the two clades (CpDV-1 and -2) in these sites is highlighted in light green. Numbers on top of the alignment (black) indicate the position in the alignment, while the numbers underneath the alignment (grey, italics) indicate the position in the *vp* gene sequence alignment used in selection analyses. The  $\eta$  and  $\alpha$  symbols represent  $3_{10}$ -helices and  $\alpha$ -helices, respectively. All  $\beta$ -strands are shown as arrows and strict  $\beta$ -turns as TT letters (ESPrpt.3).

sequences, correlation-coefficient = 0.20,  $R^2 = 0.04$ ). Thus, we further investigated whether a true clock signal existed in our data as previously described (Trovão et al. 2015). Our results proved the lack of a true clock signal in the data (Supplementary Table S4), and due to this lack, we performed a discrete phylogeographic reconstruction by fixing the evolutionary rate. For this, we analysed twenty-eight sequences from samples where both *ns1* and *vp* ORFs were sequenced (which represent 40% of the

CpDV genome) from the same individuals. We have excluded laboratory lines and recombinant sequences, and grouped the samples according to their collection site into ten geographical regions to construct the discrete phylogeographic trees (Fig. 5).

Based on the predefined geographic regions, CpDV-1 strains analysed here were most likely to be originated from Africa (Sub-Saharan and North Africa), representing 0.67 posterior probability (Fig. 5 and Supplementary Table S5). Although the uncertainty for



the origin of CpDV-2 was higher, our results suggest that it could have originated from Northern Mediterranean region (Europe and Near-East) or from Americas, representing a combined 0.48 and 0.46 posterior probability, respectively (Fig. 5 and Supplementary Table S5). Although the exact probabilities differed, these results were consistent with the results obtained from *vp* gene and *ns* gene separately (Supplementary Fig. S1).

Even though some samples collected from geographically close sites tend to group together (e.g. Tunisia samples; Fig. 5), CpDV-1 and -2 clades are found more or less everywhere without a clear geographical pattern. However, the fact that CpDV-1 clade was found in more samples than CpDV-2 could suggest the success of CpDV-1 compared with CpDV-2 in nature.

**Table 2.** Diversifying selection acting on VP on the branch between CpDV-2 and -1 clades.

Site	AA change	P
213 (4232)	Thr-Ser	<0.001
243 (4142)	Glu-Asp	0.004
290 (4001)	Thr-Ser	<0.001
311 (3938)	Leu-Ile	<0.001
324 (3899)	Met-Leu	<0.001
439 (3554)	Ile-Leu	<0.001

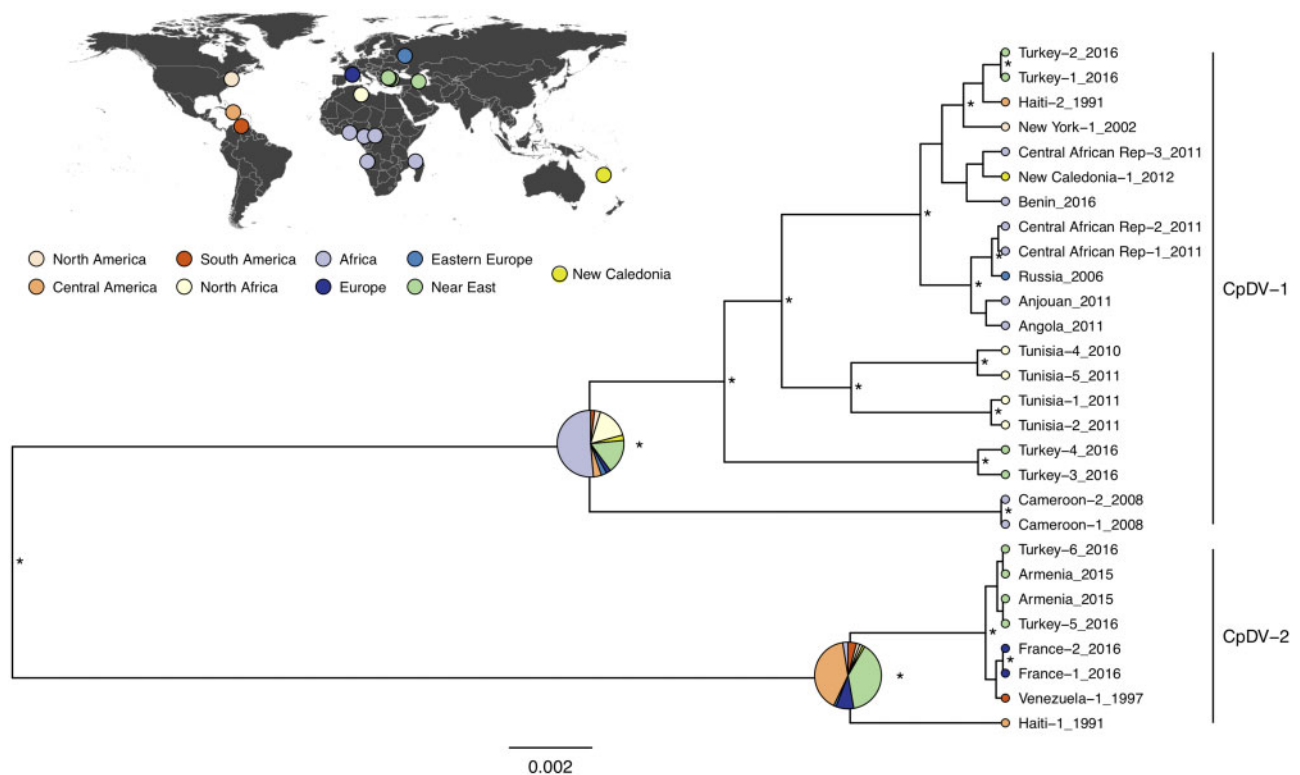
P values indicate the significance of  $\omega > 1$  for a given site. Site indicates the position of a given amino acid within the translated protein alignment, position of the first nucleotide of a given amino acid in the reference CpDV genome is given in brackets. AA, amino acid.

## 4. Discussion

### 4.1 High CpDV prevalence and diversity in nature

In this study, we showed a high CpDV prevalence (48%) in natural *C. pipiens* (s.l.) populations using a PCR diagnostic test. CpDV detection, thus, might either represent persistent infections or the detection of CpDV-related endogenised sequences. The latter seems unlikely as, in contrast to other densoviruses, CpDV-related sequences have never been found inserted in arthropod genomes so far (Liu et al. 2011; Thézé et al. 2014; Metegnier et al. 2015; François et al. 2016). Furthermore, if CpDV amplification would have resulted from endogenised sequences in germinal cells we would expect to see the amplification in all the samples, at least within a given collection site. However, it was not the case as we found CpDV-negative larvae co-existing with the CpDV-positive ones. Therefore, it seems more likely that CpDV circulates in *C. pipiens* populations, probably resulting in covert infections. Indeed, we could verify previously the presence of capsids by electron microscopy in the insectary colonies (Altinli et al. 2019). However, this was not possible for samples from natural populations, as we mostly used DNA that has been previously extracted from these samples.

Our knowledge on MDV prevalence in natural mosquito populations is limited in particular for ambidensovirus, where CpDV is the only representative in the genus. For MDVs belonging to *Brevidensovirus* genus, two studies on *Aedes aegypti* densovirus Thai strain reported prevalence in natural populations (Kittayapong, Baisley, and O'Neill 1999; Rwegoshora, Baisley, and Kittayapong 2000). Results have shown a slightly lower prevalence than CpDV, where the *Ae. aegypti* densovirus (Thai



**Figure 5.** Bayesian MCC tree from the discrete phylogeographic joint analysis of twenty-eight sequences from the genes *ns1* and *vp* of CpDV. Tree tips are coloured by sampling location. Pie charts display the locations' posterior probability for common ancestor nodes of CpDV-1 and -2. Nodes with a high posterior probability support (>0.85) are indicated with an asterisk. Scale is in nucleotide substitutions per site. The tree was rooted under the assumption of a strict, empirically fixed, molecular clock (Supplementary Material for the annotated tree file).



strain) prevalence was 6–35% in *Ae. aegypti* adults (Kittayapong, Baisley, and O'Neill 1999) and 19% in *Anopheles minimus* larvae (Rwegoshora, Baisley, and Kittayapong 2000). In this latter study, the prevalence was influenced by seasonal environmental changes and rainfall, which probably reflected variations in host population density and consequently in virus transmission rates (Kittayapong, Baisley, and O'Neill 1999; Rwegoshora, Baisley, and Kittayapong 2000). Here, we could not formally assess the influence of such environmental factors, as our sampling does not allow to determine whether prevalence can vary in a population over time. However, the analysed mosquitoes were collected from contrasting ecosystems and the presence of CpDV in half of these samples demonstrates the presence of CpDV in natural *C. pipiens* populations living in quite different environmental conditions.

Our knowledge of CpDV diversity was so far limited to one CpDV reference genome isolated from our insect facilities more than 20 years ago (i.e. Reference\_1994; Jousset, Baquerizo, and Bergoin 2000). Our results show that CpDV strains sequenced are polymorphic on *ns1* and *vp* genes. Despite the several surveys of different mosquito species including *Culex* (Zhai et al. 2008; Ma et al. 2011; Ng et al. 2011), MDV diversity described so far belong to the *Breviadensovirus* genus. We can speculate that CpDV-related virus discovery was biased due to their narrower host range. Indeed, brevidensoviruses such as *Ae. aegypti* densovirus can infect several species from the genera *Aedes*, *Culex*, and *Culiseta* and several mosquito cell lines (Carlson, Suchman, and Buchatsky 2006), while most mosquito cell lines are poorly susceptible to CpDV (Jousset, Baquerizo, and Bergoin 2000). The fact that *C. pipiens* (s.l.)-derived cell lines are less commonly used could also explain why diversity of CpDV-related viruses was unknown so far, especially given that brevidensoviruses have mostly been discovered from commonly used cell lines originated from *Aedes* and *Anopheles* mosquitoes, where they cause persistent infections (Jousset et al. 1993; O'Neill et al. 1995; Chen et al. 2004; Paterson et al. 2005; Ren, Hoiczky, and Rasgon 2008; Zhai et al. 2008).

#### 4.2 Role of selection and recombination in CpDV evolution

To date, recombination has been mostly studied for vertebrate parvoviruses (Shackelton et al. 2007; Lefeuvre et al. 2009; Martin et al. 2011). Although far less studied, intra-genus recombination has been predicted among *Ambiadensovirus* and *Iteradensovirus* genera (Francois et al. 2014; Martynova, Schal, and Mukha 2016). In this study, despite the similarity of their general topology, *ns1* and *vp* phylogenetic trees of CpDV strains exhibited some discrepancies, which can be explained by recombination. Although we confirmed *ns* and *vp* discrepancies for putative recombinant samples by re-sequencing, we cannot exclude that the putative recombination could result from an amplification bias in co-infected individuals. Nevertheless, this detection of recombination between CpDV clades suggests that viruses belonging to different clades of CpDV were replicating in the same individual and the same cells, at least at some point of their evolution. Contrarily, recombination within *ns* and *vp* sequences were scarcely detected, probably due to the compactness of the viral genome and the high selective pressures exerted on these genes as any change can be deleterious.

Our results also show that most of the sites within *ns1* and *vp* ORFs evolve under purifying selection, with only some sites evolving under episodic diversifying selection in VP. These observations are consistent with the strong purifying selection

that is generally observed in parvoviruses (Truyen et al. 2011; Martynova, Schal, and Mukha 2016; Stamenković et al. 2016). Martynova, Schal, and Mukha (2016) have reported a strong signal of purifying selection in the NS of most *Ambiadensovirus* species including CpDV and a diversifying selection signal in the VP of CpDV.

Additionally, we detected diversifying selection within a domain forming a structural motif, conserved among most parvoviruses' capsid proteins including the VP4 of GmDV. The ambiadensovirus capsid can be composed of one to four VP proteins (VP1–4) in differing ratios. VP4, the smallest and the most abundant protein constitutes the surface of the capsid and is thus crucial for cell recognition and internalisation (Simpson et al. 1998; Croizier et al. 2000; Bruemmer et al. 2005; Vendeville et al. 2009). Although the structure of CpDV capsid is yet to be determined, the cross-reaction of *Junonia coenia* densovirus (JcDV) anti-capsid antibodies with CpDV capsids suggests their structural similarities (Jousset, Baquerizo, and Bergoin 2000; Altinli et al. 2019). Four VP proteins have previously been detected in CpDV using SDS-PAGE and assigned to VP1–4, with respective sizes of 90, 64, 57, and 12 kDa (Jousset, Baquerizo, and Bergoin 2000). The latter, VP4, being too small in size to include the  $\beta$  barrel required for capsid assembly, is thought to rather result from a proteolytic cleavage of a larger VP (Jousset, Baquerizo, and Bergoin 2000; Baquerizo-Audiot et al. 2009). Although we do not know how the *vp* ORF of CpDV encodes the capsid proteins yet, the positive selection in this conserved domain might point to favourable changes in host range and specificity, as already suggested for the GmDV and JcDV (Simpson et al. 1998; Multeau et al. 2012).

#### 4.3 Phylogeography of CpDV strains

Lack of temporal signal in our data prevented us from inferring the evolutionary rates and the time to the most recent ancestor of the CpDV strains. An informative temporal signal from tips of the tree (i.e. without any internal calibration point) requires a fast evolutionary rate and/or a wide sampling time span (Pybus 2006). In the case of vertebrate parvoviruses, the evolutionary rate appears to be fast enough to perform these analyses (Truyen et al. 2011; Osterhaus et al. 2014; Stamenković et al. 2016; Fan et al. 2017), although similar analyses have never been conducted for densoviruses. Therefore, it is difficult to conclude whether the reason for this lack of temporal signal was a lower evolutionary rate of invertebrate parvoviruses compared with vertebrate parvoviruses, or a limited number of samples and/or the size of the genetic markers used in our study.

The phylogeographic reconstruction of CpDV clades suggested these two clades probably originated from different continents. This could also be pointing at CpDV origins in different species of the *C. pipiens* (s.l.) with different geographical distributions. For example, *Culex quinquefasciatus* is widespread in tropical and sub-tropical regions, including the African lowlands, whereas *C. pipiens* is widespread in Europe (Farajollahi et al. 2011). As our samples were mostly from Africa, Europe, and Near-East, more samples especially from local populations in Americas and Asia are needed to elucidate the phylogeographic history of CpDV. Overall, whether they originate from different continents or not, the low geographic structure of the different CpDV strains suggests their efficient dispersal. This could be achieved by the migration of mosquitoes of the *C. pipiens* (s.l.) by plane and boat, as either larval or adult stages (Farajollahi et al. 2011). Infected adult mosquitoes can also transmit densoviruses to new larval/oviposition sites horizontally (Wise de Valdez

et al. 2010) or to their offspring vertically (Barreau, Jousset, and Bergoin 1997; Ren et al. 2014; Altinli et al. 2019). Densoviruses are also stable in water bodies for at least a year (Buchatsky 1989), as such they could have also been carried environmentally or by human activity alone.

#### 4.4 CpDV evolution in laboratory colonies

Monitoring the CpDV evolution in laboratory conditions from 1990 to 2018, we observed a putative recombinant CpDV strain in 1997 and a swift change from CpDV-2 to -1 in the following years. This implies that during this period a CpDV-1 and a -2 strain have co-existed. The fact that CpDV-1 has not been replaced since in the laboratory line, and is more frequent in nature might also reflect that CpDV-1 clade could be more adapted to their hosts and/or more competitive. Given the importance of the capsid protein on virulence, the amino acid substitutions we observed in the VPs of CpDV-1 and -2 clades might have played a decisive role in virus fitness. Indeed, CpDV was first isolated following high larval mortality in 1994, although the CpDV which currently circulates in our laboratory *C. pipiens* lines seem to cause persistent infections with no apparent symptoms (Altinli et al. 2019).

Similarly, such persistence of CpDV might also occur in natural populations if the fitness cost of the infection is low and could explain their high prevalence in nature. Mosquito densoviruses can indeed cause persistent infections in host cells (O'Neill et al. 1995) and have negligible fitness effects on mosquitoes (Ren et al. 2014), but some highly pathogenic ones also exist (Carlson, Suchman, and Buchatsky 2006). For CpDV, more experiments are required to understand their fitness costs and determine the role, if any, played by mutations in VPs from the two clades on virulence and/or host specificity. Future studies on the evolution of insect-specific viruses with varying degrees of pathogenicity can give valuable insights on the effect of the different host-virus interactions on virus evolution and virus population dynamics in general. In this context, CpDV-*C. pipiens* system represents a valuable model to study host-virus interactions and evolution both in natural populations and with experimental approaches in laboratory conditions.

#### Data availability

CpDV sequences used in this study are available on GenBank, and their respective accession numbers can be found in the [Supplementary Table S3](#). Viral metagenomics data used for primer design is submitted on the SRA database of NCBI with the accession number PRJNA529606. All the other data used in the study are available in the [Supplementary Material](#).

#### Acknowledgements

We are grateful to two anonymous reviewers for their comments and suggestions that improved this study. We would like to thank Vincent Robert for the samples from Armenia, Benoit Assogba for the samples from Benin, Beniamino Caputo for samples from Italy, and Sandra Unal and Patrick Makoundou for technical support. We also thank Bram Vrancken and Philippe Lemey for useful discussions regarding the methodology. Sequencing data were produced at the LabEX 'Centre Méditerranéen de l'Environnement et de la Biodiversité' in the GENSEQ platform. S.L. is a postdoctoral research fellow funded by the Fonds Wetenschappelijk Onderzoek (FWO, Belgium).

**Conflict of interest:** None declared.

#### Supplementary data

Supplementary data are available at [Virus Evolution](#) online.

#### References

- Afanasiev, B. N., Galyov, E. E., and Buchatsky, L. P. (1991) 'Nucleotide Sequence and Genomic Organization of *Aedes* Densovirus', *Virology*, 336: 323–36.
- Altinli, M., Gunay, F., Alten, B. (2018) 'Wolbachia Diversity and Cytoplasmic Incompatibility Patterns in *Culex pipiens* Populations in Turkey', *Parasit Vectors*, 11: 198.
- et al. (2019) 'Sharing Cells with Wolbachia: The Transovarian Vertical Transmission of *Culex pipiens* Densovirus', *Environmental Microbiology*, 21: 3284–98.
- Atyame, C. M. et al. (2015) 'Stable Coexistence of Incompatible *Wolbachia* along a Narrow Contact Zone in Mosquito Field Populations', *Molecular Ecology*, 24: 508–21.
- Baquerizo-Audiot, E. et al. (2009) 'Structure and Expression Strategy of the Genome of *Culex pipiens* Densovirus, a Mosquito Densovirus with an Ambisense Organization', *Journal of Virology*, 83: 6863–73.
- Barreau, C., Jousset, F. X., and Bergoin, M. (1997) 'Venereal and Vertical Transmission of the *Aedes albopictus* Parvovirus in *Aedes aegypti* Mosquitoes', *The American Journal of Tropical Medicine and Hygiene*, 57: 126–31.
- Belyi, V. A., Levine, A. J., and Skalka, A. M. (2010) 'Sequences from Ancestral Single-Stranded DNA Viruses in Vertebrate Genomes: The Parvoviridae and Circoviridae Are More Than 40 to 50 Million Years Old', *Journal of Virology*, 84: 12458–62.
- Boonnak, K. et al. (2015) 'Phylogenetic Analysis Reveals Genetic Variations of Densovirus Isolated from Field Mosquitoes in Bangkok and Surrounding Regions', *The Southeast Asian Journal of Tropical Medicine and Public Health*, 46: 207–14.
- Boublik, Y., Jousset, F. X., and Bergoin, M. (1994) 'Complete Nucleotide Sequence and Genomic Organization of the *Aedes albopictus* Parvovirus (AaPV) Pathogenic for *Aedes aegypti* Larvae', *Virology*, 200: 752–63.
- Bruemmer, A. et al. (2005) 'Structure of an Insect Parvovirus (*Junonia coenia* Densovirus) Determined by Cryo-Electron Microscopy', *Journal of Molecular Biology*, 347: 791–801.
- Buchatsky, L. (1989) 'Densovirus of Bloodsucking Mosquitoes', *Diseases of Aquatic Organisms*, 6: 145–50.
- Carlson, J., Suchman, E., and Buchatsky, L. (2006) 'Densoviruses for Control and Genetic Manipulation of Mosquitoes', *Advances in Virus Research*, 68: 361–92.
- Chen, S. et al. (2004) 'Genetic, Biochemical, and Structural Characterization of a New Densovirus Isolated from a Chronically Infected *Aedes albopictus* C6/36 Cell Line', *Virology*, 318: 123–33.
- Cotmore, S. F. et al. (2019) 'ICTV Virus Taxonomy Profile: Parvoviridae', *Journal of General Virology*, 100: 367–8.
- et al. (2014) 'The Family Parvoviridae', *Archives of Virology*, 159: 1239–47.
- , and Tattersall, P. (2014) 'Parvoviruses: Small Does Not Mean Simple', *Annual Review of Virology*, 1: 517–37.
- Croizier, L. et al. (2000) 'Protein Requirements for Assembly of Virus-Like Particles of *Junonia coenia* Densovirus in Insect Cells', *Journal of General Virology*, 81: 1605–13.
- Duffy, S., Shackelton, L. A., and Holmes, E. C. (2008) 'Rates of Evolutionary Change in Viruses: Patterns and Determinants', *Nature Reviews Genetics*, 9: 267–76.

- Dumas, E. et al. (2013) 'Population Structure of Wolbachia and Cytoplasmic Introgression in a Complex of Mosquito Species', *BMC Evolutionary Biology*, 13: 181.
- Duron, O. et al. (2005) 'Transposable Element Polymorphism of Wolbachia in the Mosquito *Culex pipiens*: Evidence of Genetic Diversity, Superinfection and Recombination', *Molecular Ecology*, 14: 1561–73.
- Elena, S. F., and Sanjuán, R. (2007) 'Virus Evolution: Insights from an Experimental Approach', *Annual Review of Ecology, Evolution, and Systematics*, 38: 27–52.
- Fan, W. et al. (2017) 'Analysis of Evolutionary Processes of Species Jump in Waterfowl Parvovirus', *Frontiers in Microbiology*, 8: 421.
- Farajollahi, A. et al. (2011) 'Bird Biting' Mosquitoes and Human Disease: A Review of the Role of *Culex pipiens* Complex Mosquitoes in Epidemiology', *Infection, Genetics and Evolution*, 11: 1577–85.
- Firth, C. et al. (2009) 'Insights into the Evolutionary History of an Emerging Livestock Pathogen: Porcine Circovirus 2', *Journal of Virology*, 83: 12813–21.
- Francois, S. et al. (2014) 'A Novel Itera-Like Densovirus Isolated by Viral Metagenomics from the Sea Barley *Hordeum marinum*', *Genome Announcements*, 2: 13–4.
- François, S. et al. (2018) 'Viral Metagenomics Approaches for High-Resolution Screening of Multiplexed Arthropod and Plant Viral Communities', in V. Pantaleo, and M. Chiumenti (eds.) *BT - Viral Metagenomics: Methods and Protocols*, pp. 77–95. New York, NY: Springer New York.
- et al. (2016) 'Discovery of Parvovirus-Related Sequences in an Unexpected Broad Range of Animals', *Scientific Reports*, 6: 1–13.
- Gascuel, O., Gouy, M., and Lyon, D. (2010) 'SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building', *Molecular Biology and Evolution*, 27: 221–4.
- Georghiou, G. P., Metcalf, R. L., and Gidden, F. E. (1966) 'Carbamate-Resistance in Mosquitos', *Bulletin of the World Health Organization*, 35: 691–708.
- Green, M. R., and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual*. New York, NY: Cold Spring Harbor Laboratory Press.
- Gudenkauf, B. M. et al. (2014) 'Discovery of Urchin-Associated Densoviruses (Family Parvoviridae) in Coastal Waters of the Big Island, Hawaii', *Journal of General Virology*, 95: 652–8.
- Guillemaud, T., Pasteur, N., and Rousset, F. (1997) 'Contrasting Levels of Variability between Cytoplasmic Genomes and Incompatibility Types in the Mosquito *Culex pipiens*', *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264: 245–51.
- Hewson, I. et al. (2014) 'Densovirus Associated with Sea-Star Wasting Disease and Mass Mortality', *Proceedings of the National Academy of Sciences of the United States of America*, 111: 17278–83.
- Jousset, F. X., Baquerizo, E., and Bergoin, M. (2000) 'A New Densovirus Isolated from the Mosquito *Culex pipiens* (Diptera: Culicidae)', *Virus Research*, 67: 11–6.
- et al. (1993) 'A Parvo-Like Virus Persistently Infecting a C6/36 Clone of *Aedes albopictus* Mosquito Cell Line and Pathogenic for *Aedes aegypti* Larvae', 29: 99–114.
- Kalyanamoorthy, S. et al. (2017) 'ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates', *Nature Methods*, 14: 587–9.
- Kapoor, A., Simmonds, P., and Lipkin, W. I. (2010) 'Discovery and Characterization of Mammalian Endogenous Parvoviruses', *Journal of Virology*, 84: 12628–35.
- Kittayapong, P., Baisley, K. J., and O'Neill, S. L. (1999) 'A Mosquito Densovirus Infecting *Aedes aegypti* and *Aedes albopictus* from Thailand', *The American Journal of Tropical Medicine and Hygiene*, 61: 612–7.
- Koonin, E. V., Dolja, V. V., and Krupovic, M. (2015) 'Origins and Evolution of Viruses of Eukaryotes: The Ultimate Modularity', *Virology*, 479–480: 2–25.
- Kosakovsky Pond, S. L., and Frost, S. (2005) 'Not so Different after All: A Comparison of Methods for Detecting Amino Acid Sites under Selection', *Molecular Biology and Evolution*, 22: 1208–22.
- Krupovic, M. (2013) 'Networks of Evolutionary Interactions Underlying the Polyphyletic Origin of ssDNA Viruses', *Current Opinion in Virology*, 3: 578–86.
- Leal, E. et al. (2013) 'Interclade Recombination in Porcine Parvovirus Strains', *Journal of General Virology*, 94: 464.
- Lefeuve, P. et al. (2009) 'Widely Conserved Recombination Patterns among Single-Stranded DNA Viruses', *Journal of Virology*, 83: 2697–707.
- Lemey, P. et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5: e1000520.
- Liu, H. et al. (2011) 'Widespread Endogenization of Densoviruses and Parvoviruses in Animal and Human Genomes', *Journal of Virology*, 85: 9863–76.
- Ma, M. et al. (2011) 'Discovery of DNA Viruses in Wild-Caught Mosquitoes Using Small RNA High Throughput Sequencing', *PLoS One*, 6: e24758–7.
- Martin, D. P. et al. (2011) 'Recombination in Eukaryotic Single Stranded DNA Viruses', *Viruses*, 3: 1699–738.
- Martynova, E. U., Schal, C., and Mukha, D. V. (2016) 'Effects of Recombination on Densovirus Phylogeny', *Archives of Virology*, 161: 63–75.
- Metegnier, G. et al. (2015) 'Comparative Paleovirological Analysis of Crustaceans Identifies Multiple Widespread Viral Groups', *Mobile DNA*, 6: 1–13.
- Mietzsch, M., Pénczes, J. J., and Agbandje-Mckenna, M. (2019) 'Twenty-Five Years of Structural Parvovirology', *Viruses*, 11: 362.
- Mochizuki, T. et al. (2012) 'Archaeal Virus with Exceptional Virion Architecture and the Largest Single-Stranded DNA Genome', *Proceedings of the National Academy of Sciences of the United States of America*, 109: 13386–91.
- Multeau, C. et al. (2012) 'Four Amino Acids of an Insect Densovirus Capsid Determine Midgut Tropism and Virulence', *Journal of Virology*, 86: 5937–41.
- Murrell, B. et al. (2012) 'Detecting Individual Sites Subject to Episodic Diversifying Selection', *PLoS Genetics*, 8: e1002764.
- Ng, T. F. F. et al. (2011) 'Broad Surveys of DNA Viral Diversity Obtained through Viral Metagenomics of Mosquitoes', *PLoS One*, 6: e20579.
- O'Neill, S. L. et al. (1995) 'Insect Densoviruses May Be Widespread in Mosquito Cell Lines', *Journal of General Virology*, 76: 2067–74.
- Osterhaus, A. et al. (2014) 'Molecular Epidemiology of Seal Parvovirus, 1988–2014', *PLoS One*, 9: e112129.
- Paradis, E., and Schliep, K. (2018) 'Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R', *Bioinformatics*, 35: 526–8.
- Pasture, N., Nancé, E., and Bons, N. (2001) 'Tissue Localization of Overproduced Esterases in the Mosquito *Culex pipiens* (Diptera: Culicidae)', *Journal of Medical Entomology*, 38: 791–801.
- Paterson, A. et al. (2005) 'Mosquito Densoviruses Cause Dramatically Different Infection Phenotypes in the C6/36 *Aedes albopictus* Cell Line', *Virology*, 337: 253–61.
- Pietilä, M. K. et al. (2009) 'An ssDNA Virus Infecting Archaea: A New Lineage of Viruses with a Membrane Envelope', *Molecular Microbiology*, 72: 307–19.



- Pocquet, N. et al. (2013) 'Multiple Insecticide Resistances in the Disease Vector *Culex p. quinquefasciatus* from Western Indian Ocean', *PLoS One*, 8: e77855–14.
- Pond, S. L. K., Frost, S. D. W., and Muse, S. V. (2005) 'HyPhy: Hypothesis Testing Using Phylogenies', *Bioinformatics*, 21: 676–9.
- et al. (2006) 'Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm', *Molecular Biology and Evolution*, 23: 1891–901.
- Pybus, O. G. (2006) 'Model Selection and the Molecular Clock', *PLoS Biology*, 4: e151–8.
- Rambaut, A. et al. (2018) 'Posterior Summarisation in Bayesian Phylogenetics Using Tracer 1.7', *Systematic Biology*, 67: 901–4.
- et al. (2016) 'Exploring the Temporal Structure of Heterochronous Sequences Using TempEst (Formerly Path-O-Gen)', *Virus Evolution*, 2: vew007.
- Ren, X., Hoiczky, E., and Rasgon, J. L. (2008) 'Viral Paratransgenesis in the Malaria Vector *Anopheles gambiae*', *PLoS Pathogens*, 4: 4–11.
- et al. (2014) '*Anopheles gambiae* Densovirus (AgDENV) Has Negligible Effects on Adult Survival and Transcriptome of Its Mosquito Host', *PeerJ*, 2: e584.
- Robert, X., and Gouet, P. (2014) 'Deciphering Key Features in Protein Structures with the New ENDscript Server', *Nucleic Acids Research*, 42: 320–4.
- Rogers, S. O. and Bendich A. J. (1989) 'Extraction of DNA from Plant Tissues', in S. B. Gelvin, R. A. Schilperoort, and D. P. S. Verma (eds.) *Plant Molecular Biology Manual*, pp. 73–83. Dordrecht: Springer Netherlands.
- Rosario, K., and Breitbart, M. (2011) 'Exploring the Viral World through Metagenomics', *Current Opinion in Virology*, 1: 289–97.
- Rwegoshora, R. T., Baisley, K. J., and Kittayapong, P. (2000) 'Seasonal and Spatial Variation in Natural Densovirus Infection in *Anopheles minimus* s.l. in Thailand', *The Southeast Asian Journal of Tropical Medicine and Public Health*, 31: 3–9.
- Shackelton, L. A. et al. (2007) 'Comparative Analysis Reveals Frequent Recombination in the Parvoviruses', *Journal of General Virology*, 88: 3294–301.
- et al. (2005) 'High Rate of Viral Evolution Associated with the Emergence of Carnivore Parvovirus', *Proceedings of the National Academy of Sciences of the United States of America*, 102: 379–84.
- et al. (2006) 'Phylogenetic Evidence for the Rapid Evolution of Human B19 Erythrovirus', *Journal of Virology*, 80: 3666–9.
- Simpson, A. A. et al. (1998) 'The Structure of an Insect Parvovirus (*Galleria mellonella* Densovirus) at 3.7 Å Resolution', *Structure*, 6: 1355–67.
- Smith, M. D. et al. (2015) 'Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection', *Molecular Biology and Evolution*, 32: 1342–53.
- Stamenković, G. G. et al. (2016) 'Substitution Rate and Natural Selection in Parvovirus B19', *Scientific Reports*, 6: 1–9.
- Suchard, M. A. et al. (2018) 'Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10', *Virus Evolution*, 4: 1–5.
- Thézé, J. et al. (2014) 'Remarkable Diversity of Endogenous Viruses in a Crustacean Genome', *Genome Biology and Evolution*, 6: 2129–40.
- Tijssen, P. et al. (2016) 'Diversity of Small, Single-Stranded DNA Viruses of Invertebrates and Their Chaotic Evolutionary Past', *Journal of Invertebrate Pathology*, 140: 83–96.
- Trovão, N. S. et al. (2015) 'Host Ecology Determines the Dispersal Patterns of a Plant Virus', *Virus Evolution*, 1: vev016.
- Truyen, U. et al. (2011) 'High Rate of Viral Evolution in the Capsid Protein of Porcine Parvovirus', *Journal of General Virology*, 92: 2628–36.
- Vendeville, A. et al. (2009) 'Densovirus Infectious Pathway Requires Clathrin-Mediated Endocytosis Followed by Trafficking to the Nucleus', *Journal of Virology*, 83: 4678–89.
- Weaver, S. et al. (2018) 'Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes', *Molecular Biology and Evolution*, 35: 773–7.
- Wickham, H. (2009) *Ggplot2*. New York, NY: Springer International Publishing.
- Wise de Valdez, M. R. et al. (2010) 'A Large Scale Laboratory Cage Trial of *Aedes* Densonucleosis Virus (AeDENV)', *Journal of Medical Entomology*, 47: 392–9.
- Xu, H. et al. (2012) 'Evidence for Natural Recombination between Mink Enteritis Virus and Canine Parvovirus', *Virology Journal*, 9: 252.
- Ye, J. et al. (2012) 'Primer-BLAST: A Tool to Design Target-Specific Primers for Polymerase Chain Reaction', *BMC Bioinformatics*, 13: 134.
- Yu, G. et al. (2017) 'Ggtree: An R Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data', *Methods in Ecology and Evolution*, 8: 28–36.
- Zhai, Y. G. et al. (2008) 'Isolation and Characterization of the Full Coding Sequence of a Novel Densovirus from the Mosquito *Culex pipiens* Pallens', *Journal of General Virology*, 89: 195–9.