Virus Research 313 (2022) 198728

Contents lists available at ScienceDirect

Virus Research

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

Insect-specific viruses in the *Parvoviridae* family: Genetic lineage characterization and spatiotemporal dynamics of the recently established *Brevihamaparvovirus* genus

Paulo Morais [a,*], Nídia S. Trovão [b], Ana B. Abecasis [c], Ricardo Parreira [a]

[a] *Unidade de Microbiologia Médica, Portugal/Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical (IHMT), Universidade Nova de Lisboa (NOVA), Rua da Junqueira 100, Lisboa 1349-008, Portugal*

[b] *Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA*

[c] *Unidade de Saúde Pública Internacional e Bioestatística, Portugal/Global Health and Tropical Medicine (GHTM), Instituto de Higiene e Medicina Tropical (IHMT), Universidade Nova de Lisboa (NOVA), Lisboa, Portugal*

ARTICLE INFO

Keywords:
Brevihamaparvovirus
Parvoviridae
Virus taxonomy
Phylogenetic analysis
Spatiotemporal dynamics

ABSTRACT

The analysis of the viruses allocated to the recently established *Brevihamaparvovirus* genus (*Parvoviridae* family), which includes all previously known brevidensoviruses, has not yet been carried out on an extensive basis. As a result, no detailed genetic lineage characterization has ever been performed for this group of insect-specific viruses.

Using a wide range of molecular tools, we have explored this taxon by calculating Shannon entropy values, intra- and inter-taxon genetic distances, analysed sequence polymorphisms, and evaluated selective pressures acting on the viral genome. While the calculated *Brevihamaparvovirus* mutation rates were within the range of those of other parvoviruses, their genomes look to be under strong purifying selection, and are also characterized by low diversity and entropy. Furthermore, even though recombination events are quite common among parvoviruses, no evidence of recombination (either intra or intergenic) was found in the *Brevihamaparvovirus*es sequences analyzed. An extended taxonomic analysis and reevaluation of existing *Brevihamaparvovirus*es sequences, many still unclassified, was performed using cut-off values defining NS1 identity between viral sequences from the *Parvovirus* family. Two existing genetic lineages, Dipteran *Brevihamaparvovirus* 1 and Dipteran *Brevihamaparvovirus* 2, were rearranged and the creation of a new one, Dipteran *Brevihamaparvovirus* 3, was suggested. Finally, despite the uncertainties associated with both the time estimates of the most recent common ancestors, which could span from twenty thousand years before the current era to way earlier (in the last century), and the dispersal routes proposed for *Brevihamaparvovirus*es sequences by phylodynamic reconstruction, the analyses here presented could help define how future studies should be conducted as more isolates continue to be identified in the future, and contribute to eliminating possible analytical biases.

1. Introduction

Mosquitoes are important vectors for many pathogenic agents with (re)emerging potential, and many of these correspond to viruses (Gould et al., 2017), some of which pose threats to public health, and may cause epidemics that get considerable worldwide attention (Barzon, 2018). However, over the last decade, in addition to the discovery of many pathogenic viruses in association with hematophagous arthropods, many studies have also brought to light a plethora of so-called insect-specific viruses (ISV). The latter encompass a genetically diverse assemblage of taxonomically distinct viruses, which all share restricted/null replication capacity in vertebrate cells (Calisher and Higgs, 2018; Abudurexiti et al., 2019). Among them stand the viruses of the *Brevihamaparvovirus* genus, which belongs to the *Parvoviridae* family, which was first established in 1975 and groups viruses found in most major vertebrate and invertebrate clades (Cotmore et al., 2014).

Parvoviruses are small (23-28 nm), icosahedral-shaped, non-enveloped viral agents with single-stranded linear DNA (ssDNA) genomes ranging from 4 to 6 kilobases (kb). Two major coding regions determine the expression of non-structural (NS) and structural (VP) proteins, the

* Corresponding author.
E-mail address: a21000757@ihmt.unl.pt (P. Morais).

https://doi.org/10.1016/j.virusres.2022.198728
Received 21 January 2022; Received in revised form 2 March 2022; Accepted 3 March 2022
Available online 5 March 2022

largest of which (the so-called non-structural protein 1, or NS1), displays a highly conserved helicase superfamily domain with helicase and ATPase activity (Cotmore et al., 2019). Until 2020, parvoviruses were allocated to either the *Densovirinae* (infecting invertebrates) or *Parvovirinae* (infecting vertebrates) subfamilies, with initial subfamily demarcation exclusively supported by the topologies of phylogenetic trees (Muzyczka and Berns, 2001). However, a recent taxonomy revision took into account not only phylogenetic criteria, but also amino acid sequence similarity values calculated from comparisons of the sequences of the NS1 protein or its helicase domain (Pénzes et al., 2020). While high sequence identity for most of the NS1 protein characterized the *Parvovirinae* subfamily, the same did not apply to the *Densovirinae* subfamily. In addition, new densoviruses have also been unexpectedly isolated from vertebrates (Bochow et al., 2015; Yang et al., 2016), adding to the heterogeneous nature of this subfamily, and supporting its division into two distinct ones: the *Densovirinae* and *Hamaparvovirinae*. While hamaparvoviruses share (on average) approximately 30% of NS1 amino acid identity, they only share around 20% of sequence identity when their helicase domain is compared with that of other parvoviruses. Furthermore, as the insect-specific brevidensoviruses (*Densovirinae* subfamily) shared around 30% of NS1 protein identity with other hamaparvoviruses, they were inserted into the *Hamaparvovirinae* subfamily and the genus renamed *Brevihamaparvovirus* (Pénzes et al., 2020).

Among parvoviruses, the members of the genus *Brevihamaparvovirus* have some of the smallest ssDNA genomes (approximately 4 kb), with three open reading frames encoding two non-structural proteins (NS1, NS2) and a capsid protein (VP) (Bergoin and Tijssen, 2010). While NS1 has been known to be crucial for the initiation of viral DNA replication, NS2 participates in viral egress from the nucleus where viral replication takes place (Chen et al., 2021). The VP gene encodes a capsid protein that is essential for viral entry into host cells and the production of infectious virus (Sánchez-Martínez et al., 2012). Also, the viral genome is characterized by the presence of a unique non-coding terminal palindromic hairpin loop that is required for DNA replication (Afanasiev et al., 1991).

*Brevihamaparvovirus*es have been isolated from various mosquito cell-lines (Afanasiev et al., 1991; Ren et al., 2008) as well as wild mosquitoes, mostly from different species of *Aedes* and *Culex* from Asia (Kittayapong et al., 1999), the Americas (Sadeghi et al., 2018), Europe (Silva et al., 2019) and Africa (Morais et al., 2020), suggesting a widespread distribution. As many of these viruses have been isolated in recent years, we undertook an extensive genetic diversity analysis of this taxon using multiple bioinformatic tools. These included a comprehensive phylogenetic and an attempted spatiotemporal dispersal reconstruction analysis of these ISV.

2. Materials and methods

2.1. Dataset and sequence alignment preparation

The compilation of the different nucleotide (nt) and amino acid (aa) sequence into the datasets used in this work was based on the selection of sequences encoding non-structural protein 1 (NS1), non-structural protein 2 (NS2) and viral protein (VP) of members of the *Brevihamaparvovirus* genus, available at GenBank as of 01/08/2021. These were either directly identified via their respective accession numbers (described in previous publications), or indirectly singled out as a product of sequence similarity searches using BLASTn. All sequences available to date were downloaded, and additional information including GenBank accession number, host species, geographic origin, and collection date was also obtained. Furthermore, for comparative and phylogenetic purposes, representative datasets containing NS1 nt and aa sequences (the most conserved genomic section) of viruses from other genera in the *Parvoviridae* family were also constructed.

Multiple alignments of viral sequences in each dataset were performed using the iterative G-INS-I method as implemented in MAFFT vs. 7 (Katoh and Standley, 2013), followed by their edition using GBlocks (Castresana, 2000). The multiple sequence alignments were systematically verified to ensure the correct alignment of homologous codons using BioEdit 7.0.5 (Hall, 1999). Multiple alignments of NS1 aa sequences from multiple genus in the *Parvoviridae* family were also performed. In this case, the MAFFT iterative L-INS-I option was employed by alignment refinements using the G-INS-I method.

2.2. Assessment of the temporal and phylogenetic signals of *Brevihamaparvovirus* sequence datasets

The evolutionary information contained in all the sequence datasets compiled (phylogenetic signal) was assessed by Likelihood Mapping (Strimmer and von Haeseler, 1997) using TREE-PUZZLE v5.3 (Schmidt et al., 2002). Datasets for which $>$ 85% of the sequence quartets were totally resolved (randomly selected), were considered of good/high phylogenetic resolving power.

A visual inspection of the degree of genomic sequence divergence accumulation over the sampling time interval (i.e. temporal signal) in all nt datasets was carried out using an exploratory linear regression approach, assuming the topology obtained in a Maximum Likelihood (ML) tree, estimated under an unconstrained clock and the GTR+Γ+I substitution model using IQ-TREE (Trifinopoulos et al., 2016). Root-to-tip divergence values were plotted as a function of sampling time using the TempEst software (Rambaut et al., 2016).

2.3. Genetic diversity analyses

The estimation of genetic distance values (corrected with the Kimura-2P formula) was carried out using MEGAX (Kumar et al., 2018). Heat maps were designed based on pairwise evolutionary distances obtained using the Heatmapper webserver (Babicki et al., 2016). Visualization of genome organization for *Brevihamaparvovirus* (as well as that of selected members of other *Parvoviridae* genera) was executed using Open Reading Frame (ORF) Finder (available in https://www.ncbi.nlm.nih.gov/orffinder/), and the SMART webserver (Letunic and Bork, 2018). The presence of conserved domains in viral protein sequences was investigated using CD-Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). For protein variation analyses, single amino-acid polymorphisms (SAPs) were detected. The indicated amino acid coordinates correspond to those in the *Aedes albopictus* densovirus 2 sequence (accession number X74945).

The analyses of selective pressure acting on individual sites of codon alignments were carried out using the Single Likelihood Ancestor Counting (SLAC) and the Fixed Effects Likelihood (FEL) methods as implemented in Datamonkey (Kosakovsky Pond and Frost, 2005), or the SNAP tool (http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html), the latter exploring a simpler method for calculation of synonymous and non-synonymous substitutions (Nei and Gojoborit, 1986). Principal coordinate analyses were carried out using PCOORD (http://www.hiv.lanl. gov/content/sequence/PCOORD/PCOORD.html). Additionally, possible recombination events were investigated using the Recombination Detection Program 4 (RDP4) software (Martin et al., 2015).

2.4. Phylogenetic analyses using maximum likelihood and Bayesian approaches

To assess the relationship between the isolates belonging to the *Brevihamaparvovirus* genus and the remaining genera in the *Parvoviridae* family, phylogenetic reconstructions were carried out using NS1 aa sequences and the ML optimization criterion assuming the Whelan and Goldman (WAG) model, as defined by IQ-TREE (Trifinopoulos et al., 2016). The stability of the obtained tree topologies was assessed by bootstrapping based on 1000 re-samplings of the original sequence data. Phylogenetic reconstructions (ML) using *Brevihamaparvovirus*

ORF-specific nt and aa datasets were performed using the GTR+Γ+I and WAG model, respectively, as suggested by IQ-TREE. Once again, the stability of the obtained tree topologies was assessed by bootstrapping with 1000 re-samplings of the original sequence data. All phylogenetic reconstructions were carried out assuming a relaxed uncorrelated lognormal molecular clock model (Ho et al., 2005) as indicated by the ML Clock Test implemented in MEGA X, allowing for the accommodation of among-lineage rate variation.

Time-calibrated phylogenetic and phylogeographic histories were obtained using a Bayesian statistical framework, as implemented in the BEAST v1.10 software package (Suchard et al., 2018), and using the GTR+Γ+I model. To investigate the sensitivity of the estimate for the time to the Most Recent Common Ancestor (tMRCA) concerning the coalescent priors used, the performance of constant, exponential, logistic, and expansion parametric population demographic growth priors (Drummond et al., 2003; Griffiths and Tavaré, 1994) was tested against that of nonparametric ones, including the Bayesian Gaussian Markov Random Field (GMRF) Skyride (Minin et al., 2008), Skygrid (Gill et al., 2013) and Skyline (Drummond et al., 2005). This preliminary comparative analysis was carried out using all the VP dataset sequences available. Bayes factor (BF) support for predictors was calculated using marginal likelihood estimates (MLE) (inferred using Path Sampling (PS) and Stepping-Stone (SS) approaches) for each candidate model, and then comparing the ratio of the marginal likelihood estimates for the set of candidate models being compared.

A minimum number of two, and up to a maximum of twenty, independent Markov chain Monte-Carlo (MCMC) runs were performed using BEAST v1.10 until 1-3 $\times$ 10^8 states were sampled, with at least 10% of which being discarded as burn-in. The length of the MCMC analyses was defined as a function of chain convergence which was followed using the Tracer software v1.7.1 (http://beast.bio.ed.ac.uk/tracer). The latter was also used to check for adequate effective sample size (ESS) higher than 200 after the removal of burn-in. The trees were logged on every 10,000th MCMC step, and the trees distribution was summarized using the TreeAnnotator software v1.8.3 as a maximum clade credibility (MCC) tree, with median heights as the node heights in the tree. The FigTree v1.4.2 software was used to visualize the phylogenetic trees (http://tree.bio.ed.ac.uk/software/figtree/).

2.5. Continuous phylogeography

The geographic spread of *Brevihamaparvovirus*es in continuous space was studied using a phylogenetic Brownian diffusion approach that models the change in geographic coordinates (latitude and longitude) along each branch in the phylogenetic reconstruction (Lemey et al., 2010). As an alternative to the latter, relaxed random walk (RRW) extensions that model branch-specific variation in dispersal rates similar to uncorrelated relaxed clock approaches was also used (Drummond et al., 2006). The assessment of BF support for the diffusion priors was calculated using MLE as described above for the coalescent demographic priors.

The spatiotemporal reconstruction of the spread *Brevihamaparvovirus* was visualized on the Spatial Phylogenetic Reconstruction of Evolutionary Dynamics software (SpreaD3; Bielejec et al., 2016), using a custom-made geoJSON world map (https://geojson-maps.ash.ms/).

3. Results

3.1. Comparative genomic coding architecture and genetic diversity analyses

Public genomic database mining allowed the creation of three datasets containing *Brevihamaparvovirus* (BHP) nt and aa sequences. These included 60 NS1 sequences, 31 NS2 sequences, and 40 VP sequences (additional information available on Supplementary Table 1). Most sequences (~90%) were originally identified in association with specimens of either *Culex sp.* or *Aedes sp.* mosquitoes, with the remaining ones being amplified from *Anopheles, Culiseta, Armigeres*, and *Haemagogus*. Five of the BHP sequences were obtained from C6/36 cell cultures and two from chronically infected cell lines (Aag2 and SuaB5). Additionally, for phylogenetic and other comparative analyses, NS1 aa sequences were also compiled for viruses from each genus/species in all *Parvoviridae* subfamilies (Supplementary Table 2).

Brevihamaparvovirus ORF organization, as seen in Fig. 1, is almost identically shared between all BHP, with two distinct regions coding the non-structural proteins (NS1 and NS2) and one viral structural protein (VP). However, the genome of one BHP (accession number MH188047), isolated in 2016 from a *Culex* mosquito, displayed two ORFs that encoded structural proteins (Fig. 1). This genetic organization recalls that of typical *Ambidensovirus* (*Densovirinae* subfamily) sequences (Supplementary Fig. 1). The composition of ORF-coding sequences looked similar between different genera in the *Hamaparvovirinae* subfamily, with the only noticeable exception found among the members of the *Ichthamaparvovirus* genus, for which an NS2 coding sequence could not be identified. On the contrary, ORF organization inside the *Densovirinae* and *Parvovirinae* subfamilies was found to be quite disparate. An alternative ORF-coding sequence, encoding the so-called assembly-activating protein (AAP), which promotes capsid assembly (Earley et al., 2017), seemed to exist only in the genome of the members of the *Dependoparvovirus* genus.

Overall, comparison of all individual BHP ORF-coding sequences disclosed low mean genetic distances, with the lowest value associated with the NS1 protein (0.053), followed by NS2 (0.064) and VP (0.077). Pairwise evolutionary distances (PEDs) were calculated between all BHP sequences and analyzed using a heatmap (Supplementary Fig. 2). PEDs analysis did provide insights into possible segregation of three different groups of sequences inside the *Brevihamaparvovirus* genus, with those inside each group sharing low PEDs values, while slightly higher values were observed when sequence comparisons extended towards those from other viral groups. To compare genetic distance values between members of the different *Parvoviridae* subfamilies, overall mean genetic distances were calculated individually (for each subfamily) for the most conserved ORF-coding region (NS1), using datasets containing sequences from each genus/species. The inclusion of a more divergent group of NS1-coding sequences into a single dataset naturally raised the average genetic distance values of the *Hamaparvovirinae, Densovirinae* and *Parvovirinae* subfamilies to 0.499, 0.502 and 0.503, respectively, with no apparent significant difference between all values.

Shannon entropy is a quantitative measure of uncertainty in a dataset of nucleotide or amino acid sequences, and it may be considered as a measure of variation in DNA and protein sequence alignments for assessment of genetic diversity in a cross-sectional sense. When applied to the analysis of BHP sequences, Shannon entropy calculations showed low values for all BHP ORF-coding sequences. NS1-coding region did show slightly lower entropy values in most amino acids (Supplementary Fig. 3A), which should be expected due to its key role in viral DNA replication. In addition, no substantial differences in entropy values were found between different subfamilies in the *Parvoviridae* families when evaluating NS1 aa sequences entropy (Supplementary Fig. 3B).

3.2. Phylogenetic signal, selective pressure, impact of genetic recombination, and sequence divergence accumulation throughout time

In order to assess the extent to which selective pressure and/or intra/intergenic recombination events could impact phylogenetic reconstructions, both were metrics evaluated using specific bioinformatic tools. No evidence of either intra or intergenic recombination events were detected for either full-length genomes or any of the genes analysed, using all detection methods on the RDP4 software. Estimation of omega (ω) values (corresponding to the ratio of non-synonymous to synonymous substitutions) was performed for BHP using three different methods (SLAC, FEL, and SNAP). These analyses were carried out for all

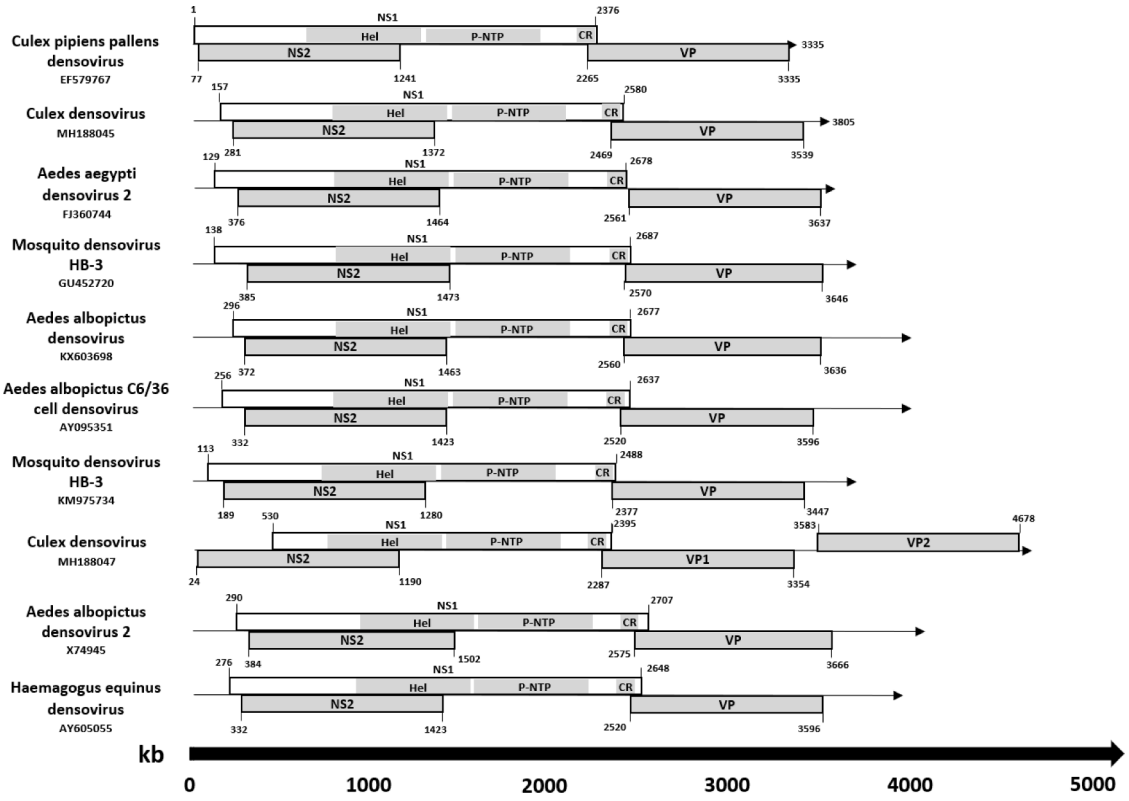

Fig. 1. Schematic representation of nucleotide sequences for ten different *Brevihamaparvovirus*, with different ORFs identified; NS – non-structural protein; VP – viral protein; Hel – Helicase; P-ATP – P-loop NTPase; CR – coiled region.

ORF-coding regions (Supplementary Table 3), and the results obtained indicated that the whole genome seems to be under purifying selection, as deduced by overall low ω values, especially in the case of the VP-coding region (p-value < 0.05). Site-specific selection analysis also revealed little to no evidence of positively selected sites. Only 2 codons in the NS2 gene were identified as evolving under diversifying selection, and even so, this observation was only supported by one analysis methods used (FEL).

Since no evidence of recombination events or positive selective pressure acting on the BHP genome were detected (which could compromise phylogenetic reconstruction), the phylogenetic signals of each nt dataset were evaluated using likelihood mapping (Table 1). The obtained results showed high phylogenetic resolving power for both the NS1 and VP sequence datasets, with 88.1% and 85.5%, of totally resolved randomly selected 10,000 quartet replicates respectively. The NS2 gene showed slightly lower phylogenetic resolving power with 77.0% of totally resolved sequence quartets. These results indicated that phylogenetic reconstructions based on the analysis of alignments of any viral ORFs (with the possible exception of NS2), would produce unambiguous phylogenetic trees.

To assess the extent to which all BHP sequence datasets contained detectable signals indicating expanding sequence divergence throughout time, a standard linear regression exploration of root-to-tip genetic distances as a function of sampling time was performed. Only the NS1-coding region did not reveal clear evidence for a substantial temporal signal (Table 1), even after the removal of outlier sequences that could have a negative impact on temporal signal assessment. Nevertheless, while for both the NS2 and VP sequence datasets, a substantial temporal signal was found, we selected the VP gene as the prime candidate for continuous phylogeography analysis (see below) as it also displayed the highest phylogenetic signal. However, both the very narrow temporal date sampling interval (with an average date-range of 20 years) and the *a priori* unknown average rates of evolutionary change for *Brevihamaparvovirus*es, could influence temporal signal assessment. As far as the latter was concerned, nucleotide substitution rates were estimated using the sequences of the BHP VP gene, while assuming a relaxed molecular clock model (Drummond et al., 2006). This was supported by ML test of the molecular clock hypothesis, which systematically rejected the null hypothesis of equal nucleotide substitution rates along the branches of the trees (Supplementary Table 4A). Substitution rate values varied depending on the coalescent priors used and ranged from 1.16×10^{-3} to 2.24×10^{-4} substitutions per site/per year.

Table 1
Phylogenetic signal (as assessed by likelihood mapping) and root-to-tip (sequence divergence as a function of time) of *Brevihamaparvovirus* sequences using datasets of all ORF-coding sequences.

Likelihood Mapping	NS1	NS2	VP
Totally resolved quartets	88.1%	77.0%	85.5%
Partially resolved quartets	3.5%	0.7%	1.8%
Unresolved quartets	8.4%	22.3%	12.7%
Root-to-tip analysis (r^2)	0.028	0.220	0.450

3.3. Phylogenetic analyses

Previous reports have stated that NS1 proteins of viruses belonging to the same genus share between 35-40% of amino acid sequence identity with a minimum shared query cover of 80% (between any two members being compared), while simultaneously clustering as a robust monophyletic lineage (Pénzes et al., 2020). Accordingly, a phylogenetic reconstruction of the evolutionary relationships within the *Parvoviridae* family (Supplementary Fig. 4) and the genetic distance values indicated in Supplementary Table 5 clearly show the singularity of the *Brevihamaparvovirus* genus. While the analysis of NS1 sequences suggested that *Brevihamaparvovirus*es share common ancestry with the members of *Parvovirinae* subfamily, this shared ancestry was not supported by

bootstrap analysis.

Considering the above mentioned, (i) high phylogenetic signal of NS1 and VP sequence datasets, and (ii) the absence of evidence for intra/intergenic recombination, (iii) or of positive selection acting as a driver of virus evolution, the evolutionary relationships between only BHP were investigated using ML phylogenetic tree reconstruction (Supplementary Fig. 5). No substantial differences were found between the NS1 and VP ML trees, and in both the BHP sequences were segregated into three distinct monophyletic clades. Furthermore, when the previously defined minimum of 85% of identity (based on the analysis of NS1 aa sequences) was considered to unite *Brevihamaparvovirus*es as members of the same species (Pénzes et al., 2020), the different BHP genetic lineages did seem to correspond to distinct viral species, consistent with the NS1 tree topology. However, not only have many new, and therefore unclassified, sequences been described recently from the Americas (Sadeghi et al., 2018), Asia (Fu et al., 2017), Europe, and Africa (Morais et al., 2020; Silva et al., 2019), a dissent was observed between the NS1 tree topology and the currently accepted taxonomic assignments (Fig. 2A). For these reasons, Fig. 2B indicates a suggested correction of the BHP genetic lineage assignment, confirming the previously defined Dipteran *Brevihamaparvovirus* (DB) 1 and DB2 genetic lineages, and suggesting the establishment of a third one, designated DB3. The existence of the DB1-3 genetic sublineages was also supported by PCOORD analysis (Fig. 2C), as well as by the NS1 genetic diversity grouping indicated by heatmaps (Supplementary Fig. 2).

Multiple single amino-acid polymorphisms were identified when NS1 aa variation was analyzed taking into account sequence comparisons between different *Brevihamaparvovirus* branches, further supporting the identified viral sublineages (Fig. 2C). The analysis of lineage-specific SAPs indicated four of them to be exclusively found on DB2 (479D, 487S, 524A and 546H) and DB3 (74H, 496D, 539K and 586S) sequences, while polymorphisms 481R and 522K characterized the DB1 lineage.

3.4. Continuous phylogeography

In an attempt to infer the population dynamics of BHP sequence dispersal through space and time, using a dataset with high phylogenetic resolving power and reasonable temporal signal (VP; r^2=0.45), we first tested the performance of parametric demographic priors against that of non-parametric ones. The obtained results clearly showed that the non-parametric priors consistently performed better than the parametric ones, indicated by both Bayes factor (Supplementary Table 4A) and effective sample size (ESS) values, which were consistently higher for non-parametric priors. The obtained results also pointed towards the Bayesian Skyline as the coalescent prior of choice, as judged by marginal likelihood and ESS values (consistently higher than 200). A comparative assessment of the performance of a strict Brownian vs. several RRW diffusion models for BHP was also performed, allowing us to evaluate what would be the best geographic diffusion model to be used for spatiotemporal dispersal analysis. The obtained MLE values (shown in Supplementary Table 4B) suggested that a Gamma-RRW prior was the one best fit to explain its dispersal dynamics.

Results for the spatiotemporal analysis were summarized as a MCC phylogenetic tree (Supplementary Fig. 6), as well as projected into maps using the SpreaD3 software (Fig. 3). High Highest Probability Density (HPD) intervals for MRCA ages were estimated for almost all nodes, especially for the MRCA for all BHP sequences. While these high HPD intervals appear to dissipate as the estimates moved towards the more recent nodes, the analysis of root age dates should be interpreted with caution. Although it seems clear that two viral lineages have diverged well in the past into two distinct clades, evaluation of dispersion routes between BHP's oldest and also the most recent ancestors were not firmly supported by our analyses. With the available data, our analysis suggests the possibility of an expansion of one of the viral lineages into two distinct ones dating more than two thousand years ago (95% HDP: (-10263)-2014) in both eastern and western directions (Asia and North America). The other expansion event, for which a Middle Eastern

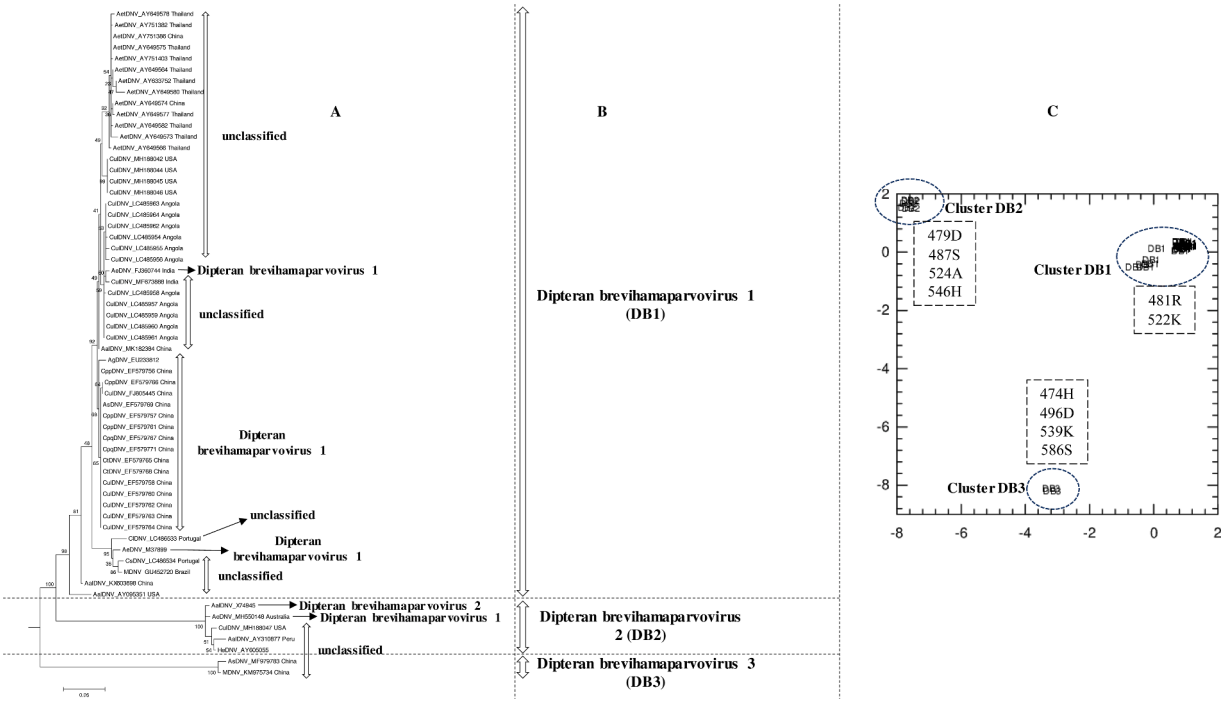

Fig. 2. Maximum likelihood tree of NS1 *Brevihamaparvovirus* sequences juxtaposed to (A) the current ICTV taxonomy information and (B) a classification scheme of parvoviruses combining the tree topology and the species-defining NS1 identity percentages (as previously defined by Pénzes et al. (2020); (C) Principal coordinate analysis carried out for NS1 coding sequences, where each sequence is identified by the major branch they belong to (as shown in Fig. 1B). Single amino-acid polymorphisms exclusive to each identifiable group are indicated in association with the DB1-3 clusters. The indicated positions relate to those of the Aedes albopictus densovirus 2 sequence (accession number X74945).

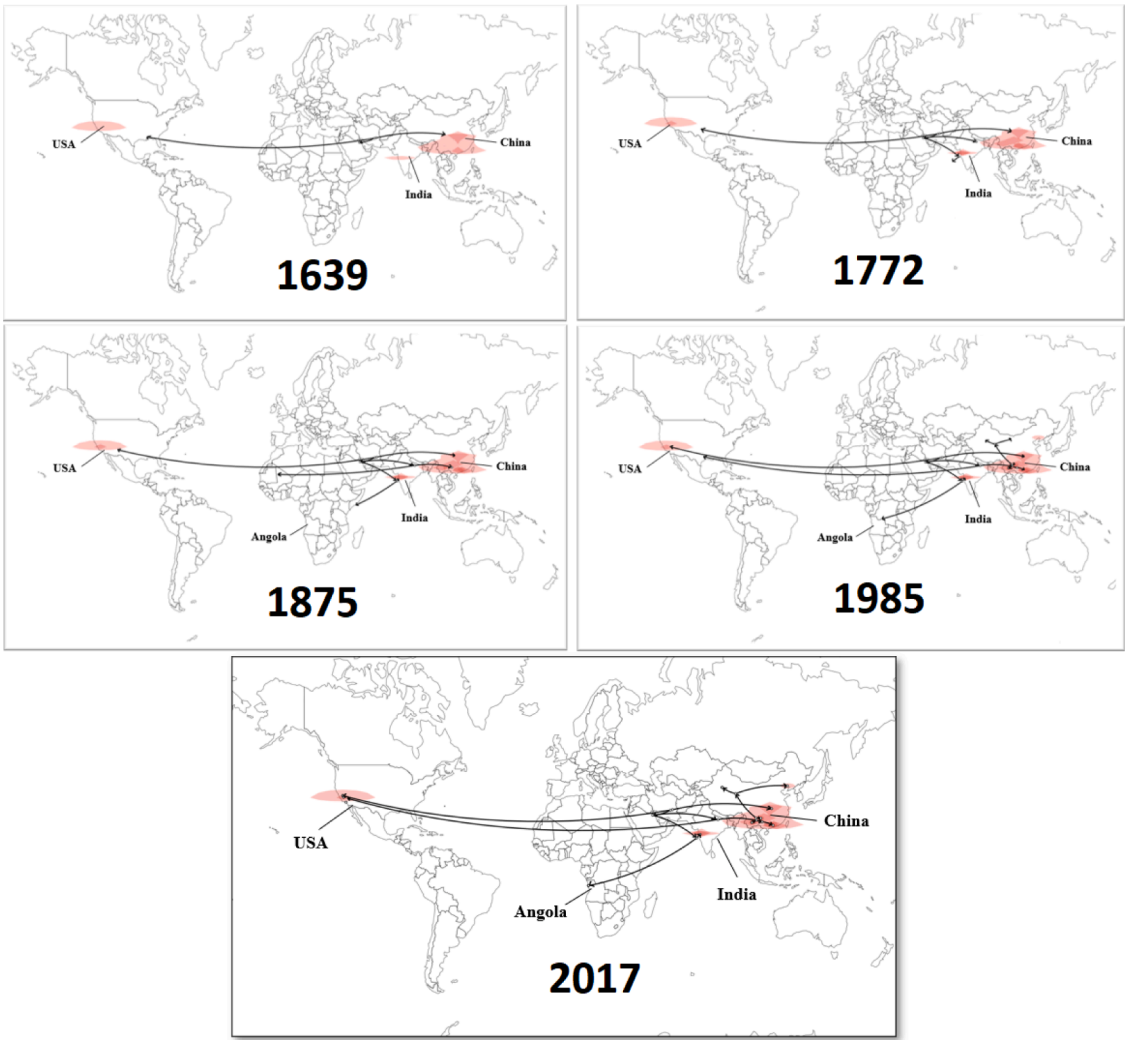

Fig. 3. Spatiotemporal reconstruction of *Brevihamaparvovirus* spread visualized on SpreaD3 software, based on the MCC tree represented in Supplementary Fig. 6.

(possibly Iranian) geographical origin was suggested (though not statistically supported), seemed to have split into two different routes in the 1700s. More recent years marked the expansion of one of the possible Indian clades into Africa (Angola) in the early 2010s (95% HPD: 1990-2016, and strongly supported with a posterior probability of 1), while the other clade seemed to have expanded in two totally different directions. These included movements towards Asia, starting in the early 1870 (95% HPD: 1068-1995) as well as North America, starting as late as in 2014 (95% HPD: 1995-2016). Both dispersal routes were strongly supported by location posterior probability values (of 0.8 and 1, respectively).

4. Discussion

Unlike the case of most genera in the *Parvoviridae* family that join viruses identified in association with both vertebrate and invertebrate species (Pénzes et al., 2020), the members of the *Brevihamaparvovirus* genus have only been, up to the present day, found in mosquitoes. To what extent is this association with Diptera absolute is still open to debate. In fact, the analysis of other ISVs has shown that in specific cases they eventually bypass host range restrictions imposed by certain host cells (Junglen et al., 2017), thus expanding their host range (Morais et al., 2021). These observations open the possibility that among the large diversity of viruses associated with viral *taxa* whose members are supposedly restricted to replicate in insect cells, some may acquire the capacity to adapt to a larger collection of hosts. Such a case has already been described in the *Parvoviridae* family, when sequences of members of the *Ambidensovirus* genus (*Densovirinae* subfamily), mostly associated with insect hosts until recently, have been recently detected in vertebrate hosts such as ducks (accession number MW306771) and cranes (accession number MW046535).

As it has been previously considered for other ISVs (Goenaga et al., 2020), *Brevihamaparvovirus*es are not only widespread in a variety of wild mosquito species, but have also been found to interfere *in vitro* with the replication of *bona fide* arboviruses, such as reducing the severity of the cytopathic effects induced by dengue virus infection in C6/36 cells persistently infected with Aedes albopictus *Brevihamaparvovirus* (Burivong et al., 2004). These observations highlight the potential use of certain arboviruses as biological agents to interfere with vector competence. These viruses seem to be able to integrate their genomes into that of infected cells, therefore they can also be exploited as vehicles for stable expression of heterologous proteins in insect cells (Öhlund et al., 2019). All these facts justify a more detailed approach to gather information on this specific genus in the *Parvoviridae* family, where extensive genetic research has been scarce. In fact, studies addressing the analysis of members of the *Brevihamaparvovirus* genus (formerly designated as *Brevidensovirus*) have been almost non-existent, with only sporadic reports on the detection of a new viral genome, the analysis of its genetic structure (Chen et al., 2004), or what phenotypic effects are associated with their replication on specific cell lines (Paterson et al.,

2005). Broader genetic studies have focused exclusively on either the taxonomy revision of the *Parvoviridae* family as a whole (Pénzes et al., 2020) or the characterization of the phylogenetic relationships of its members (Cotmore et al., 2019), with sporadic reports of estimation of nucleotide substitution rates and selective pressure analysis for some parvovirus (Stamenković et al., 2016). However, to this date, the members of the *Brevihamaparvovirus* genus have not been analysed in detail, including assessments of genetic diversity, selective pressure, Shannon entropy, or spatiotemporal dynamics. In this regard, this study provides new insights into the genetic characteristics of BHP, as well as what evolutionary events may have contributed to their dissemination, and what technical aspects limit our ability to describe it precisely and with detail.

Our genetic analyses were based on the assembly of multiple datasets of sequences from three different genomic regions (NS1, NS2 and VP). The NS proteins share the majority of conserved domains and a coiled region essential for DNA replication (Bergoin and Tijssen, 2010). The regions encoding NS1 and NS2 overlap, but each protein is encoded from a distinct reading frame after alternative splicing (Chen et al., 2021). While two of the regions displayed high phylogenetic signal (NS1 and VP), only one revealed a strong temporal signal (VP-coding region) required for spatiotemporal dispersal analysis. The genomic regions with the lowest and highest overall mean genetic diversity were the NS1 and VP regions, respectively. These results confirm previous observations made in association with the study of other parvoviruses (Kapoor et al., 2010; Lu et al., 2020), for which mean genetic distance values were consistently higher for coding regions of structural proteins when comparing them to non-structural proteins. Recombination events, which seem to commonly affect the evolution parvoviruses of equine (Lu et al., 2020), geese and Muscovy ducks (Shen et al., 2015), or even humans (Khamrin et al., 2013) seem to affect the VP region in particular. However, no recombination events seemed to have affected the evolution of the analysed BHP. Since the generation of recombinant genomes requires co-/superinfection events to occur, whether the apparent lack of impact of recombination on viral evolution is a direct consequence of restricted replication only in insects (mosquitoes in particular) is open to debate.

Positive selection pressure is among the factors that affect virus evolution. In the context of this study, its analysis indicated that most of the BHP genome evolves under strong purifying selection, as seen by the accumulation of a surplus of synonymous substitutions relative to the non-synonymous substitutions. Only two codons in NS2 may be under diversifying selection, but this observation was not confirmed by all method of analysis used. Furthermore, analysis of Shannon entropy, used as a measure of variation in DNA/protein sequence alignments, revealed low values for all genomic regions. Similar results have been found in canine parvovirus and human parvovirus (Shackelton et al., 2005; Stamenković et al. 2016). The biological relevance of these observations should, however, be considered with caution as they may vary significantly depending on the number of BHP sequences available (Añez et al., 2011), and a considerable number of these have been described recently (Morais et al., 2020; Silva et al., 2019). Further research regarding the clarification of what may be the mechanisms of natural selection affecting BHP is important. Indeed, reports of positive selective pressure acting on selected parvovirus genomic sites (especially in capsid protein coding regions) have been strongly connected to their ability to adsorb to new host cells (Hueffer et al., 2003a), allowing possible early detection of future BHP host-switching. Furthermore, whereas the calculated BHP substitution rates are similar to those of other parvoviruses (Shackelton et al., 2005, calculated for canine parvoviruses, ranging from 2.7×10^{-3} to 9.4×10^{-5} substitutions per site/per year; Stamenković et al., 2016, calculated for human parvovirus B19, ranging from 1.03×10^{-4} to 2.32×10^{-4} substitutions per site/per year), they are high due to the small size of the viral genome and its high replication turnover (Koppelman et al., 2007), which could influence future processes of natural selections by mutation fixation.

Considering (i) the cut-off value of NS1 sequence identity that seems suitable to define independently evolving genetic lineages for BHP, (ii) the congruence between the topologies of phylogenetic trees, (iii) the recently described BHP sequences that remained unclassified, (iv) with one of them (MH550148) being previously misclassified, our analysis did confirm the establishment of two species (Dipteran *Brevihamaparvovirus* 1 and 2) and suggests that a new one (Dipteran *Brevihamaparvovirus* 3) should be considered. This suggestion was also supported by PCOORD analysis as well as by the identification of NS1 species-specific aa polymorphisms.

Addressing phylodynamic analysis in a BHP genetic characterization study could provide relevant information regarding the estimation of the time and geographic origin of most recent common ancestor (MRCA) of all BHP sequences analysed, as well as what may have been the routes these viruses explore to spread through space and time. However, a prior assessment of genetic information contained in viral sequences is vital, since not only weak temporal signals could negatively impact the calculations of the mean MRCA time estimates (or tMRCA; Trovao et al., 2015), but the use of adequate candidate models in phylogeographic analyses is of paramount importance. Except for NS2, both NS1 and VP coding sequences showed sufficient temporal signals, which seem to be a common observation when studying parvoviruses (Stamenković et al. 2016). However, unlike observations stated in previous reports where the spatiotemporal dynamics of parvovirus was investigated considering constant or logistic population size priors, in thus study we formally demonstrate that non-parametric coalescent priors often perform better than parametric ones (Morais et al., 2021). Therefore, using adequately selected coalescent and demographic dispersal priors, our results suggested a scenario where the MRCA of the BHP under analysis may have emerged up to twenty thousand years before the current era. However, given the large 95% HDP intervals estimated for internal nodes, the proposed ages for mean tMRCA, despite giving an indication of the times of divergence, are not accurate and should be interpreted with caution. The analysis of BHP VP sequences revealed a tree root in the Middle East (with low statistical support) from where two viral lineages diverged. Despite the wide host-range and transmission routes different parvoviruses have explored to ensure their natural maintenance, in reports regarding the analysis of canine parvovirus (Giraldo-Ramirez et al., 2020), the estimated average tMRCA was 1979 (with 95% HPD range of 38–44 years), and the spread of these viruses seems to have happened quite recently. Therefore, given the very wide 95% HPD intervals associated with the older branches of the BHP phylogeography tree, the early BHP expansion events could have occurred considerably later than our analysis suggested. However, this discrepancy might be due to the use of inappropriate priors during the phylodynamic reconstruction. Multiple factors, such as massive tourism and global commercial trading, and the limited success of most vector-control programs could have a strong impact in the recent expansion of BHP. Despite its limitations, this study did provide information that could help define how future studies should be conducted, as more BHP sequences are identified.

Few reports delving into the evolutionary events that shaped the evolution of parvoviruses have been carried out to this day and, not surprisingly, these usually address pathogenic viruses. To date, none of these studies had specifically focused on the members of the *Brevihamaparvovirus* genus, which stand unique among parvoviruses thanks to their host-restriction. However, host-switching looks to be quite common in parvovirus (Hueffer et al., 2003b), so as new BHP sequences are identified in the near future, new research is crucial in order to identify how and when putative host-switching events might have eventually occurred.

Supporting information

Supplementary Fig. 1: Schematic representation of nucleotide sequences from each genus in the *Parvoviridae* family with different ORFs

identified. Not representative of the size of each ORF, only their organization and sequence. NS – non-structural protein; VP – viral protein; NP – nucleoprotein; AAP – assembly activating protein.

Supplementary Fig. 2: Heat map representing inter-sequence genetic diversity of *Brevihamaparvovirus*. Representative tree obtained on IQ-TREE (maximum likelihood, GTR+Γ+I model) based on NS1 nt sequences (reported in Supplementary Table 1), and Z-Scores estimated based on pairwise evolutionary distances using MegaX.

Supplementary Fig. 3: (A) Entropy on the basis of the Shannon function (Shannon entropy-one) for different ORF-coding sequences of *Brevihamaparvovirus*; (B) Entropy on the basis of Shannon function (Shannon entropy-one) for NS1-coding sequences of different subfamilies in the *Parvoviridae* family.

Supplementary Fig. 4: NS1 maximum likelihood phylogenetic tree of several parvovirus genera and subfamilies, estimated under a WAG substitution model using IQ-TREE (phylogeny test with 1000 bootstrap replications). Isolates are shown in Supplementary Table 2.

Supplementary Fig. 5: Maximum likelihood tree of *Brevihamaparvovirus* NS1 and VP nucleotide sequences, estimated under a GTR+Γ+I substitution model using IQ-TREE (phylogeny test with 1000 bootstrap replications). The different genetic lineages (DB1-3) are indicated.

Supplementary Fig. 6: Continuous phylogeographic analysis of *Brevihamaparvovirus* VP coding sequence. At certain nodes of the MCC tree, the geographic origin and/or the date of MRCA are indicated, with the 95% HPD values for the date of the MRCA being displayed between brackets. Posterior probability (PP) values >0.70 (for the tree topology) are indicated by circles, while the decimals associated with certain nodes indicate the inferred location PP.

CRediT authorship contribution statement

Paulo Morais: Data curation, Formal analysis, Visualization. **Nídia S. Trovão:** Writing – review & editing. **Ana B. Abecasis:** Writing – review & editing. **Ricardo Parreira:** Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work received financial support from the Global Health and Tropical Medicine Center, which is funded through FCT contract UID/Multi/04413/2013. The opinions expressed in this article are those of the authors and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2022.198728.

References

Abudurexiti, A., Adkins, S., Alioto, D., Alkhovsky, S.V., Avšič-Županc, T., Ballinger, M.J., Kuhn, J.H., 2019. Taxonomy of the order bunyavirales: update 2019. Arch. Virol. 164 (7), 1949–1965. https://doi.org/10.1007/s00705-019-04253-6.

Afanasiev, B.N., Galyov, E.E., Buchatsky, L.P., Kozlov, Y.V., 1991. Nucleotide sequence and genornic organization of aedes densonucleosis virus. Virology 185 (1), 323–336. https://doi.org/10.1016/0042-6822(91)90780-F.

Añez, G., Morales-Betoulle, M.E., Rios, M., 2011. Circulation of different lineages of dengue virus type 2 in Central America, their evolutionary time-scale and selection pressure analysis. PLoS One 6 (11). https://doi.org/10.1371/journal.pone.0027459.

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., Wishart, D.S., 2016. Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. 44 (W1), W147–W153.

Barzon, L., 2018. Ongoing and emerging arbovirus threats in Europe. J. Clin. Virol. 107, 38–47. https://doi.org/10.1016/j.jcv.2018.08.007.

Bergoin M. and Tijssen P. 2010. Densoviruses: a highly diverse group of arthropod parvoviruses, p. 57–90. *In* Asgari N. and Johnson K. N. (ed.), *Insect virology*. Horizon Scientific Press, Norwich, United Kingdom.

Bielejec, F., Baele, G., Vrancken, B., Suchard, M.A., Rambaut, A., Lemey, P., 2016. SpreaD3: interactive visualization of spatiotemporal history and trait evolutionary processes. Mol. Biol. Evol. 33, 2167–2169. https://doi.org/10.1093/molbev/msw082.

Bochow, S., Condon, K., Elliman, J., Owens, L., 2015. First complete genome of an ambidensovirus; cherax quadricarinatus densovirus, from freshwater crayfish cherax quadricarinatus. Mar. Geonomics 24, 305–312. https://doi.org/10.1016/j.margen.2015.07.009.

Burivong, P., Pattanakitsakul, S.N., Thongrungkiat, S., Malasit, P., Flegel, T.W., 2004. Markedly reduced severity of dengue virus infection in mosquito cell cultures persistently infected with aedes albopictus densovirus (AalDNV). Virology 329 (2), 261–269. https://doi.org/10.1016/j.virol.2004.08.032.

Calisher, C.H., Higgs, S., 2018. The discovery of arthropod-specific viruses in hematophagous arthropods: an open door to understanding the mechanisms of arbovirus and arthropod evolution? Annu. Rev. Entomol. 63, 87–103. https://doi.org/10.1146/annurev-ento-020117-043033.

Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.

Chen, S., Cheng, L., Zhang, Q., Lin, W., Lu, X., Brannan, J., Zhang, J., 2004. Genetic, biochemical, and structural characterization of a new densovirus isolated from a chronically infected aedes albopictus C6/36 cell line. Virology 318 (1), 123–133. https://doi.org/10.1016/j.virol.2003.09.013.

Chen, S., Miao, B., Chen, N., Chen, C., Shao, T., Zhang, X., Tong, D., 2021. SYNCRIP facilitates porcine parvovirus viral DNA replication through the alternative splicing of NS1 mRNA to promote NS2 mRNA formation. Vet. Res. 52 (1), 1–15. https://doi.org/10.1186/s13567-021-00938-6.

Cotmore, S.F., Agbandje-McKenna, M., Chiorini, J.A., Mukha, D.V., Pintel, D.J., Qiu, J., Davison, A.J., 2014. The family *Parvoviridae*. Arch. Virol 159 (5), 1239–1247. https://doi.org/10.1007/s00705-013-1914-1.

Cotmore, S.F., Agbandje-McKenna, M., Canuti, M., Chiorini, J.A., Eis-Hubinger, A.M., Hughes, J., Harrach, B., 2019. ICTV virus taxonomy profile: *Parvoviridae*. J. Gen. Virol. 100 (3), 367–368. https://doi.org/10.1099/jgv.0.001212.

Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R., Rodrigo, A.G., 2003. Measurably evolving populations. Trends Ecol. Evol. 18, 481–488.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192. https://doi.org/10.1093/molbev/msi103.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, 699–710. https://doi.org/10.1371/journal.pbio.0040088.

Earley, L.F., Powers, J.M., Adachi, K., Baumgart, J.T., Meyer, N.L., Xie, Q., Nakai, H., 2017. Adeno-associated virus (AAV) assembly-activating protein is not an essential requirement for capsid assembly of AAV serotypes 4, 5, and 11. J. Virol. 91 (3) https://doi.org/10.1128/jvi.01980-16.

Fu, S., Song, S., Liu, H., Li, Y., Li, X., Gao, X., Liang, G., 2017. ZIKA virus isolated from mosquitoes: a field and laboratory investigation in China, 2016 Sci. China Life Sci. 60 (12), 1364–1371. https://doi.org/10.1007/s11427-017-9196-8.

Gould, E., Pettersson, J., Higgs, S., Charrel, R., de Lamballerie, X., 2017. Emerging arboviruses: why today? One Health 4, 1–13. https://doi.org/10.1016/j.onehlt.2017.06.001.

Gill, M.S., Lemey, P., Faria, N.R., Rambaut, A., Shapiro, B., Suchard, M.A., 2013. Improving bayesian population dynamics inference: a coalescent-based model for multiple loci. Mol. Biol. Evol. 30, 713–724. https://doi.org/10.1093/molbev/mss265.

Giraldo-Ramirez, S., Rendon-Marin, S., Ruiz-Saenz, J., 2020. Phylogenetic, evolutionary and structural analysis of canine parvovirus (CPV-2) antigenic variants circulating in Colombia. Viruses 12 (5), 1–15. https://doi.org/10.3390/v12050500.

Goenaga, S., Goenaga, J., Boaglio, E.R., Enria, D.A., Levis, S.D.C., 2020. Superinfection exclusion studies using west nile virus and culex flavivirus strains from Argentina. Mem. Inst. Oswaldo Cruz. 115 (5), 1–5. https://doi.org/10.1590/0074-02760200012.

Griffiths, R.C., Tavaré, S., 1994. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344, 403–410.

Hall, T.A., 1999. BioEdit: a user-friendlybiological sequence alignment editor and analysis program for windows 95/98/NT.pdf. Nucleic Acids Symp. Ser. 41, 95–98.

Ho, S.Y.W., Phillips, M.J., Drummond, A.J., Cooper, A., 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. Mol. Biol. Evol. 22, 1355–1363.

Hueffer, K., Govindasamy, L., Agbandje-McKenna, M., Parrish, C.R., 2003a. Combinations of two capsid regions controlling canine host range determine canine transferrin receptor binding by canine and feline parvoviruses. J. Virol. 77 (18), 10099–10105. https://doi.org/10.1128/jvi.77.18.10099-10105.2003.

Hueffer, K., Parker, J.S.L., Weichert, W.S., Geisel, R.E., Sgro, J.Y., Parrish, C.R., 2003b. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. J. Virol. 77 (3), 1718–1726. https://doi.org/10.1128/jvi.77.3.1718-1726.2003.

Junglen, S., Korries, M., Grasse, W., Wieseler, J., Kopp, A., Hermanns, K., León-Juárez, M., Drosten, C., Kümmerer, B.M., 2017. Host range restriction of insect-

specific flaviviruses occurs at several levels of the viral life cycle. mSphere 2, 1–15. https://doi.org/10.1128/msphere.00375-16.
Kapoor, A., Simmonds, P., Slikas, E., Li, L., Bodhidatta, L., Sethabutr, O., Delwart, E., 2010. Human bocaviruses are highly diverse, dispersed, recombination prone, and prevalent in enteric infections. J. Infect. Dis. 201 (11), 1633–1643. https://doi.org/10.1086/652416.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Khamrin, P., Okitsu, S., Ushijima, H., Maneekarn, N., 2013. Complete genome sequence analysis of novel human bocavirus reveals genetic recombination between human bocavirus 2 and human bocavirus 4. Infect. Genet. Evol. 17, 132–136. https://doi.org/10.1016/j.meegid.2013.03.040.
Kittayapong, P., Baisley, K.J., O'Neill, S.L, 1999. A mosquito densovirus infecting aedes aegypti and aedes albopictus from Thailand. Am. J. Trop. Med. Hyg. 61 (4), 612–617. https://doi.org/10.4269/ajtmh.1999.61.612.
Koppelman, M.H.G.M., Rood, I.G.H., Fryer, J.F., Baylis, S.A., Cuypers, H.T.M., 2007. Parvovirus B19 genotypes 1 and 2 detection with real-time polymerase chain reaction assays. Vox Sang. 93 (3), 208–215. https://doi.org/10.1111/j.1423-0410.2007.00957.x.
Kosakovsky.Pond, S.L., Frost, S.D.W., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533.
Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549.
Lemey, P., Rambaut, A., Welch, J.J., Suchard, M.A., 2010. Phylogeography takes a relaxed random walk-in continuous space and time. Mol. Biol. Evol. 27, 1877–1885. https://doi.org/10.1093/molbev/msq067.
Letunic, I., Bork, P., 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 46 (D1), D493–D496.
Lu, G., Wu, L., Ou, J., Li, S., 2020. Equine parvovirus-hepatitis in China: characterization of its genetic diversity and evidence for natural recombination events between the Chinese and American strains. Front. Vet. Sci. 7 (March), 1–8. https://doi.org/10.3389/fvets.2020.00121.
Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1 (1), 1–5.
Minin, V.N., Bloomquist, E.W., Suchard, M.A., 2008. Smooth skyride through a rough skyline: bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25, 1459–1471. https://doi.org/10.1093/molbev/msn090.
Morais, P., Pinto, J., Jorge, C.P., Troco, A.D., Fortes, F., Sousa, C.A., Parreira, R., 2020. Insect-specific flaviviruses and densoviruses, suggested to have been transmitted vertically, found in mosquitoes collected in Angola: genome detection and phylogenetic characterization of viral sequences. Infect. Genet. Evol. 80, 104191 https://doi.org/10.1016/j.meegid.2020.104191. January.
Morais, P., Trovão, N.S., Abecasis, A.B., Parreira, R., 2021. Genetic lineage characterization and spatiotemporal dynamics of classical insect-specific flaviviruses: outcomes and limitations. Virus Res., 198507 https://doi.org/10.1016/j.virusres.2021.198507.
Muzyczka, N., Berns, K.I., 2001. *Parvoviridae*: the viruses and their replication. Fields Virol. 4 (69), 2327–2360.
Nei, M., Gojoborit, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.
Öhlund, P., Lundén, H., Blomström, A.L., 2019. Insect-specific virus evolution and potential effects on vector competence. Virus Genes 55 (2), 127–137. https://doi.org/10.1007/s11262-018-01629-9.
Paterson, A., Robinson, E., Suchman, E., Afanasiev, B., Carlson, J., 2005. Mosquito densonucleosis viruses cause dramatically different infection phenotypes in the C6/36 aedes albopictus cell line. Virology 337 (2), 253–261. https://doi.org/10.1016/j.virol.2005.04.037.
Pénzes, J.J., Söderlund-Venermo, M., Canuti, M., Eis-Hübinger, A.M., Hughes, J., Cotmore, S.F., Harrach, B., 2020. Reorganizing the family *Parvoviridae*: a revised taxonomy independent of the canonical approach based on host association. Arch. Virol. 165 (9), 2133–2146. https://doi.org/10.1007/s00705-020-04632-4.
Rambaut, A., Lam, T.T., Carvalho, L.M., Pybus, O.G., 2016. Exploring the temporal structure of heterochronous sequences using tempest (formerly path-O-gen). Virus Evol. 2, 1–7.
Ren, X., Hoiczyk, E., Rasgon, J.L., 2008. Viral paratransgenesis in the malaria vector anopheles gambiae. PLoS Pathog. 4 (8), 4–11. https://doi.org/10.1371/journal.ppat.1000135.
Sadeghi, M., Altan, E., Deng, X., Barker, C.M., Fang, Y., Coffey, L.L., Delwart, E., 2018. Virome of >12 thousand culex mosquitoes from throughout California. Virology 523 (April), 74–88. https://doi.org/10.1016/j.virol.2018.07.029.
Sánchez-Martínez, C., Grueso, E., Carroll, M., Rommelaere, J., Almendral, J.M., 2012. Essential role of the unordered VP2 n-terminal domain of the parvovirus MVM capsid in nuclear assembly and endosomal enlargement of the virion fivefold channel for cell entry. Virology 432 (1), 45–56. https://doi.org/10.1016/j.virol.2012.05.025.
Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502–504.
Shackelton, L.A., Parrish, C.R., Truyen, U., Holmes, E.C., 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. Proc. Nat. Acad. Sci. U.S.A. 102 (2), 379–384. https://doi.org/10.1073/pnas.0406765102.
Shen, H., Zhang, W., Wang, H., Zhou, Y., Shao, S., 2015. Identification of recombination between muscovy duck parvovirus and goose parvovirus structural protein genes. Arch. Virol. 160 (10), 2617–2621. https://doi.org/10.1007/s00705-015-2541-9.
Silva, M., Morais, P., Maia, C., de Sousa, C.B., de Almeida, A.P.G., Parreira, R., 2019. A diverse assemblage of RNA and DNA viruses found in mosquitoes collected in southern Portugal. Virus Res. 274, 197769 https://doi.org/10.1016/j.virusres.2019.197769. July.
Stamenković, G.G., Ćirković, V.S., Šiljić, M.M., Blagojević, J.V., Knežević, A.M., Joksić, I. D., Stanojević, M.P., 2016. Substitution rate and natural selection in parvovirus B19. Sci. Rep. 6, 1–9. https://doi.org/10.1038/srep35759. October.
Strimmer, K., von.Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. Proc. Natl. Acad. Sci. U. S. A. 94, 6815–6819.
Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4, 1–5.
Trifinopoulos, J., Nguyen, L., Haeseler, A. Von-Minh, B.Q., 2016. W-IQ-TREE : a fast online phylogenetic tool for maximum likelihood analysis 44, 232–235. https://doi:10.1093/nar/gkw256.
Trovao, N.S., Baele, G., Vrancken, B., Bielejec, F., Suchard, M.A., Fargette, D., Lemey, P., 2015. Host ecology determines the dispersal patterns of a plant virus. Virus Evol. 1 (1) https://doi.org/10.1093/ve/vev016.
Yang, W.T., Shi, S.H., Jiang, Y.L., Zhao, L., Chen, H.L., Huang, K.Y., Wang, C.F., 2016. Genetic characterization of a densovirus isolated from great tit (parus major) in China. Infect. Genet. Evol. 41, 107–112. https://doi.org/10.1016/j.meegid.2016.03.035.