

RESEARCH

Open Access



# The genomic diversity of arthropod-specific viruses reinforces the continental distribution pattern of *Aedes aegypti*

Weimar D. Briñez<sup>1,3</sup>, Daniel Alfonso Urrea<sup>2</sup>, Marina Muñoz<sup>3</sup>, Luz H. Patiño<sup>1</sup> and Juan David Ramírez<sup>1,4\*</sup>

## Abstract

**Background** *Aedes aegypti* is the primary vector of arboviruses worldwide, including dengue, Zika, chikungunya, and yellow fever. It is believed to have originated in Africa and migrated to the Americas during the sixteenth and seventeenth centuries, subsequently spreading to Asia and Oceania between the nineteenth and twentieth centuries. These mosquitoes harbor insect-specific viruses (ISVs), which represent the majority of their core virome. Phasivirus phasiense (PCLV), cell-fusing agent virus (CFAV), and *Aedes* anphevirus (AeAV) stand out for their global distribution in vectors. Within this framework, this study aims to evaluate whether ISVs can provide insights into the continental dispersal history of *Ae. aegypti*.

**Methods** A total of 96 complete sequences from three ISVs (CFAV, AeAV, and PCLV) were analyzed. These were obtained from wild *Ae. aegypti* collected across four regions: the Americas (45), Asia (38), Africa (12), and Oceania (1), as well as from laboratory colonies originally derived from wild mosquitoes. Of the 45 sequences from the Americas, 12 were newly assembled for this study from mosquito samples collected in Ibagué, Colombia in 2021 (4 per ISV). To reconstruct the vector's dispersal history and compare evolutionary patterns between viruses with and without evidence of recombination, multiple methodological approaches were used: (1) phylogenetic analyses with BEAST1 to estimate divergence times, (2) statistical tests for recombination ( $\Phi$ -test), (3) recombination network construction using SplitsTree and RDP, and (4) principal component analysis (PCA) to evaluate population structure.

**Results** The analysis of three ISVs in global *Ae. aegypti* populations allowed for the inference of the vector's historical dispersal. The results revealed: (1) genetically structured diversity patterns associated with geography, (2) evidence of recombination in PCLV, but not in AeAV and CFAV, and (3) contrasting temporal estimates suggesting multiple introductions into the Americas between the seventeenth and nineteenth centuries, as well as recent dispersal into Oceania.

**Conclusions** ISVs are promising tools for studying the dispersal and evolution of *Ae. aegypti*, although their viral dynamics can influence their effectiveness as evolutionary markers. CFAV, with recombination evidence, reflects strong connectivity among populations. AeAV, despite lacking recombination but high variability, provides accurate insights into recent dispersal. PCLV, with low diversity and regional recombination, is useful for analyzing local population structures.

**Keywords** Phylogeography, Insect vector, Recombination, Arboviruses, Biological evolution

\*Correspondence:

Juan David Ramírez

jr Ramirezgonzalez@usf.edu

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

It is estimated that *Aedes aegypti* originated during a period of drought in North Africa around 4000–6000 years ago, in what is now the Sahara Desert [1, 2]. This desertification process likely facilitated the “domestication” of *Ae. aegypti*, as human water storage practices provided artificial breeding sites, while the proximity to these human populations offered a stable and accessible blood source [2–4]. Approximately 500 years ago, *Ae. aegypti* likely arrived in the New World through the transatlantic slave trade and later expanded into Asia and the Pacific Islands, including Australia [5].

To support this hypothesis, several methodologies have been used to reconstruct the phylogeographic history of *Ae. aegypti*, including genomic analyses, single-nucleotide polymorphism (SNP), and microsatellite studies [5–8], as well as mitochondrial DNA analyses to infer patterns of divergence and dispersal [6, 9–11]. These approaches are generally useful for determining the evolution of population structure at regional scales and over long periods [12]. However, new techniques are being implemented to analyze these evolutionary processes on a local scale and over shorter time frames. These approaches are primarily aimed at understanding the biology, ecology, and population dynamics of *Ae. aegypti*, which in turn informs our understanding of the emergence and transmission of the arboviruses it vectors, such as dengue (DENV), Zika (ZIKV), chikungunya (CHIKV), and yellow fever (YFV) [13]. In this context, outbreak dynamics are studied by integrating spatial, temporal, and viral genetic data, along with information on human and vector mobility. Following this approach, it may be possible to use circulating viruses to infer the phylogeography and distribution of *Ae. aegypti* [12].

In this regard, insect-specific viruses (ISVs) are exclusive to insects and do not require a vertebrate host to complete their life cycle [14, 15], meaning that their presence in *Ae. aegypti* is generally stable and not influenced by external epidemiological factors [12]. ISVs have maintained long-term associations with their vectors, with interactions ranging from neutral coexistence to possible effects on mosquito biology and arbovirus dynamics [16]. Their persistence in *Ae. aegypti* populations has been reported worldwide [15], and they experience lower selective pressure compared with arboviruses [16], which may result in more stable evolutionary rates and a lower probability of mutational saturation [17].

Overall, ISVs represent approximately 73.8% of the mosquito virome [18]. Among them, Phasi Charoen-like virus (PCLV), from the family Phenuiviridae, constitutes between 60% and 98% of the *Ae. aegypti* core virome globally [18–20]. Likewise, cell-fusing agent virus (CFAV), the first ISV identified in 1970, belongs to the

Flaviviridae family, along with DENV, ZIKV, and YFV. Its 11 kb genome encodes a single polyprotein, and its abundance in the core virome of *Ae. aegypti* ranges between 2% and 11% [17–19, 21]. Another relevant ISV is *Aedes anphevirus* (AeAV), from the Ximoviridae family, which has an 11 kb genome and a core virome abundance ranging from 0.04% to 0.14% [22–24]. PCLV, CFAV, and AeAV are widely distributed across all continents, suggesting a stable and long-standing relationship with their host [15, 18, 19, 21, 22].

Given that ISVs have coevolved with *Ae. aegypti*, their analysis can represent a promising tool for reconstructing the evolutionary history and dispersal routes of the vector, complementing traditional genetic studies [22, 25, 26]. In this sense, although ISVs have been widely detected in *Ae. aegypti*, their intraspecific genetic variability, recombination processes, and evolutionary relationships among them, given the worldwide dispersal of the vector, have not been explored on a regional scale, nor has their potential as markers of *Ae. aegypti* dispersal. To fill this gap, our study evaluates ISVs from a genomic approach, analyzing their genetic variability across five continents, assessing whether ISVs can reflect the phylogeographic patterns of *Ae. aegypti*, offering a novel perspective that complements traditional genetic methods. To this end, we analyzed a total of 96 complete genomes from three ISV species—PCLV (L segment), AeAV, and CFAV—comprising 12 novel sequences generated for this study from wild *Ae. aegypti* collected in Ibagué, Colombia, and 84 publicly available sequences retrieved from the National Center for Biotechnology Information (NCBI).

## Methods

### Sample collection and RNA sequencing

During 2021, *Ae. aegypti* populations were collected in urban areas of the municipality of Ibagué, Tolima, Colombia (coordinates: 4°26′16″ N, 75°12′02″ W). Four pools of 50 females each were obtained from different districts within the municipality. To identify viral diversity in these populations, RNA extraction was performed for each pool using mechanical maceration with a mortar and liquid nitrogen as the homogenization method. This technique was selected owing to its advantages in preserving RNA integrity, avoiding fragmentation compared with other homogenization methods [27, 28].

For RNA extraction, the TRI Reagent® protocol (Thermo Fisher Scientific) was used [28], adding 1.5 mL to each pool to maximize RNA yield and quality. Sample quality was assessed to ensure suitability for sequencing using the 2100 Bioanalyzer; the electropherogram was examined to record band sizes associated with the 18S- and 28S-rRNA peaks as a measure of RNA integrity.

Library preparation was performed using the TruSeq Stranded Total RNA LT Sample Prep Kit (Gold) with rRNA depletion (Ribo-Zero H/M/R Gold). Sequencing was carried out on the Illumina NovaSeq 6000 platform, generating 100-base pair (bp) paired-end reads, with an average of 5.9 Gb of data for each of the four pools of 50 females processed, for a total of approximately 26 Gb of information.

### Bioinformatic processing

The four sequencing products were processed separately following the bioinformatic approach described below: Quality control was performed using FastQC version 0.11.9 [29]. Adapter sequences, low-quality bases (<Q30), and short reads (<70 bp) were removed using Prinseq-lite version 0.20.4 [30] and Trimmomatic version 0.39 [31].

Subsequently, to analyze the virome, we first filtered out the mosquito transcriptome to reduce the number of reads to be analyzed. For this, each of the cleaned sequencing products was aligned using Hisat2 [32] against the *Ae. aegypti* reference genome available from NCBI and VectorBase [33] (Chromosomes [Chr] 1–3 and Mitochondrial: NC\_035108.1, NC\_035107.1, NC\_035109.1, NC\_035159.1). Reads that did not align to the vector genome were stored in a separate FASTQ file for each pool and individually assembled de novo using Trinity version 2.13 [34] with the following parameters: \$Trinity-seqType fq-max\_memory 240G-left 1.fastq.gz-right 2.fastq.gz-CPU 32.

Assembly metrics were analyzed using descriptive statistics with the script count\_fasta\_cnsg.pl [35]. This process aimed to obtain contigs that represented partial or complete viral genomes present in the samples.

The assembled contigs were compared via Basic Local Alignment Search Tool-nucleotide (BLASTn) against the nucleotide (nt) and nr (non redundant) NCBI databases locally [36], applying a stringent threshold with an  $E \leq 1 \times 10^{-6}$  and a maximum of 20 hits per query. This produced a list of candidate viruses present in each pool.

On the basis of the viruses identified in this list, reference genomes were downloaded from the NCBI database, selecting for each virus the corresponding official reference genome. Alignments for each virus per pool were performed using Hisat2 to confirm viral presence, coverage, and depth. Concordant reads for each virus per pool were de novo assembled using Trinity, applying the same parameters and criteria mentioned above. Once the list of contigs and their taxonomic assignments was generated, alignments, organization, and orientation of the contigs into a genome or pseudomolecule were performed using the Abacas script version 1.3.1 [37] on the basis of the consulted reference genomes.

These organized viral genomes were annotated using the automatic annotation transfer tool Rapid Annotation Transfer Tools (RATT) [38], using annotated reference genomes in EMBL format, and finally verified manually. Genome coverage and depth were evaluated using Samtools version 1.9 [39] and Qualimap [40]. The reads were subsequently deposited in the NCBI database under project ID PRJNA1257366.

### Identification of genomic sequences of *Aedes anphevirus*, *Phasivirus phasiense*, and cell-fusing agent virus

Given the complete genome recovery of PCLV, CFAV, and AeAV from each of the pools obtained through sequencing and bioinformatic processing of the samples collected in Ibagué, a dataset was constructed by supplementing with sequences available in the NCBI database (<https://www.ncbi.nlm.nih.gov/>) for each of these viruses.

For sequence selection, the reference genomes of PCLV (ASM281483v1) [41], CFAV (NC\_001564.2) [42], and AeAV (NC\_075274.1) [22] were used. A BLASTn analysis was conducted using the MegaBLAST algorithm, with an  $E$ -value threshold of  $\leq 1 \times 10^{-6}$  and a maximum of 5000 matches. All other parameters were kept at their default settings.

From these results, only samples with a minimum coverage of 91% relative to the reference genome were selected [43, 44] (Supplementary Table S1, “Coverage %” column). This threshold was chosen to ensure adequate representation of the viral genome while retaining a sufficient number of sequences for comparative analyses. This practice is standard in viral genomics, where strict 100% coverage cutoffs are rarely applied owing to the inherent variability of viral quasispecies and technical limitations that can create stochastic gaps in coverage [45]. Given the high mutation rates and frequent sequence variability in RNA viruses, a strict 100% coverage cutoff would have excluded many informative sequences, especially those derived from wild-caught samples. The 91% threshold thus provides a balance between data completeness and sample inclusion, a strategy consistent with methodologies used in large-scale viral surveillance and metagenomics studies [46], minimizing biases due to missing data while preserving geographic and viral lineage diversity. Subsequently, we included sequences obtained from wild *Ae. aegypti* collected across different continents and, when available, sequences from laboratory colonies originally established from wild populations of the same regions. The distribution of sequences by region and ISV was as follows: Africa (PCLV, 7; CFAV, 2; AeAV, 3), Americas (PCLV, 20; CFAV, 12; AeAV, 13), Asia (PCLV, 26; CFAV, 7; AeAV, 5), and Oceania (AeAV, 1). For PCLV, only the L segment was selected owing to its higher

conservation and availability of sequences (Supplementary Table S1).

To ensure data quality, all reads were verified to originate from Illumina platforms or platforms with comparable Phred quality [47], minimizing sequencing errors. In the case of previously published sequences, the associated bioinformatic methodology was reviewed to rule out potential errors in data processing (Supplementary Table S1, “Technology” and “Assembly” columns). Once the database was compiled, each sequence was downloaded and classified by viral species, host, continent, country, locality, NCBI ID, percentage of coverage, sequencing technology, year of collection, genome size, and assembly technology (Supplementary Table S1).

### Alignment and phylogenomic analysis

To validate the data quality, their taxonomic identity, and their geographical relationships and associations, a phylogenetic analysis was carried out as follows: An alignment was performed using MAFFT version 7 software [48] with the FFT-NS-2 algorithm and default parameters. Subsequently, UTR regions and low-coverage areas were manually removed using UGENE version 33.0 [49], to ensure a proper comparison at the sequence level. These alignments were then saved in FASTA format.

On the basis of the alignment of each virus, a phylogenetic analysis was conducted using the maximum likelihood (ML) method on the IQ-TREE version 2 web server [50], employing the nucleotide substitution model automatically selected by ModelFinder [51]. The robustness of the topology was assessed with 10,000 bootstrap replicates [13]. A midpoint rooting was subsequently performed, and nodes with bootstrap values  $\geq 70$  were highlighted [21, 50, 51]. Finally, the geographic origin of the samples at the continental scale was color-coded using the iTOL version 6 web server [52].

### Genetic variability analysis

Subsequently, on the basis of the same alignments used for the phylogenomic analysis, viral genetic diversity was assessed using DnaSP version 6 software by applying the DNA polymorphism algorithm. The completeness of the region was verified, and nucleotide diversity ( $\pi$ ) was calculated. This index measures the average number of nucleotide differences per site between pairs of sequences in the dataset, providing an estimate of genetic variation. In addition, owing to the high mutation rate characteristic of RNA viruses, the Jukes and Cantor correction was applied to account for the possibility of multiple substitutions at the same site, which may otherwise lead to underestimation of variability [53].

Watterson’s Theta ( $\Theta_w$ ) and neutrality tests, including Tajima’s  $D$ , were also calculated to evaluate patterns

of molecular evolution and to test for deviations from neutrality, such as population expansion or selection. The resulting data were organized into a table for further comparison. Standard deviations of Watterson’s Theta ( $\Theta_w$ ), both per site and per sequence ( $\sigma^2 < \text{sub} > \theta < / \text{sub} >$ ), were then computed using the corresponding formula:

$$SD = \sqrt{\theta_w}$$

where:

SD = standard deviation

$\Theta_w$  = central estimated value of divergence

Subsequently, a dissimilarity matrix (genetic distance) was generated using RStudio on the basis of the Jukes–Cantor algorithm, from the aligned sequences for each virus. This matrix was then used in RStudio to perform a principal component analysis (PCA) on the basis of the genetic distance matrix between sequences to identify clustering patterns. For the PCA analysis, the R libraries ape, adegenet, and ggplot2 were used.

### Population structure inference and recombination detection

A phylogenetic network was constructed for each alignment using default parameters in SplitsTree version 4 [54] to visualize complex evolutionary relationships and potential recombination events. In addition, the phi test, as implemented in SplitsTree version 4, was applied to confirm signals of recombination. Furthermore, a recombination analysis was performed using RDP version 5 [55], implementing multiple algorithms, including RDP, BootScan, MaxChi, Chimaera, SiScan, and 3Seq. Only recombination events with a significance threshold of  $P < 0.05$  in at least three methods were considered [56].

### Bayesian inference and estimation of divergence times

To estimate the divergence times of PCLV, AeAV, and CFAV, as well as potential associations between their geographic distribution and the evolutionary history of the vector, a Bayesian phylogenetic inference was conducted using BEAST1 version 2.6.7 [57]. A relaxed log-normal molecular clock model was employed to allow for variable evolutionary rates across lineages. Temporal calibration of the tree was based on the collection dates of the sequences, spanning from 2002 to 2021, enabling estimation of the time to the most recent common ancestor (TMRCA). Model convergence was assessed through effective sample size (ESS) values greater than 200, ensuring robust parameter estimates.

A Bayesian skyline coalescent tree prior was used to flexibly reconstruct changes in viral population sizes over time. The nucleotide substitution model GTR+ $\Gamma$ 4+I was selected via ModelFinder [51], and an evolutionary

rate of  $1 \times 10^{-3}$  substitutions/site/year, based on previous studies of similar RNA viruses, was applied [58]. All other parameters were inferred from the dataset. To evaluate spatiotemporal dispersion patterns, sequences were grouped by continent (America, Asia, Africa, and Oceania), allowing for the inference of global dispersal events.

Markov chain Monte Carlo (MCMC) chains were run for 100 million generations, with sampling every 10,000 iterations. Parameter convergence was assessed using Tracer version 1.7.1 [59], ensuring effective sample sizes (ESS) greater than 200. A 10% burn-in was applied, along with the options `-limit 0.95` and `-hpd2D 0.95`, before summarizing the maximum clade credibility tree using TreeAnnotator version 2.6.7. Divergence times and the temporal structure of viral lineages were visualized using FigTree version 1.4.4, with each geographic origin color-coded. In addition, the time axis was set in reverse chronological order (negative direction).

#### Estimation of standard deviation (SD) from the 95% highest posterior density (HPD) interval

To estimate the standard deviation (SD) of the age of a node on the basis of the 95% highest posterior density (HPD) interval, the following approximation was used, assuming a normal distribution. Since approximately 95% of the values in a normal distribution lie within  $\pm 2$  standard deviations ( $2\sigma$ ) of the mean, the width of the HPD interval can be approximated as four times the standard deviation.

Therefore, the standard deviation is estimated using the following formula:

$$SD = \frac{HDP_{sub} - HDP_{inf}}{4}$$

where:

SD = standard deviation

HPD<sub>sup</sub> = upper limit of the 95% highest posterior density (HPD) interval

HPD<sub>inf</sub> = lower limit of the 95% HPD interval

## Results

### Dataset

Sequences were retrieved from the NCBI database through BLASTn searches using the complete genomes of each studied virus as reference. Applying strict quality criteria, sequences derived primarily from wild *Ae. aegypti* samples with genome coverage  $\geq 91\%$  were selected; in some cases, sequences from laboratory colonies originating from wild populations were also included [43, 44]. In the specific case of PCLV, the analysis focused on the L segment owing to its higher availability of complete sequences.

Four complete sequences were generated in this study for each of the three viruses—PCLV, AeAV, and CFAV—from *Ae. aegypti* individuals collected in Ibagué, Colombia in 2021. These sequences were taxonomically assigned using BLASTn against the NCBI nt database and showed high genome coverage: 99% for PCLV, 96% for AeAV, and 98% for CFAV (Fig. 1 and Supplementary Table S1). In addition, publicly available sequences were retrieved for comparative purposes, resulting in a total dataset of 53 sequences for PCLV, 22 for AeAV, and 21 for CFAV. This filtering ensured that subsequent analyses were based on complete and comparable genomic data.

### Phylogenomic analysis

To validate the taxonomy, data quality, and evolutionary relationships among the evaluated viruses, an initial phylogenetic analysis was performed using the sequences obtained in this study alongside the reference genomes for the three viruses: AeAV (NC\_075274.1), PCLV (NC\_038262.1), and CFAV (NC\_001564.2).

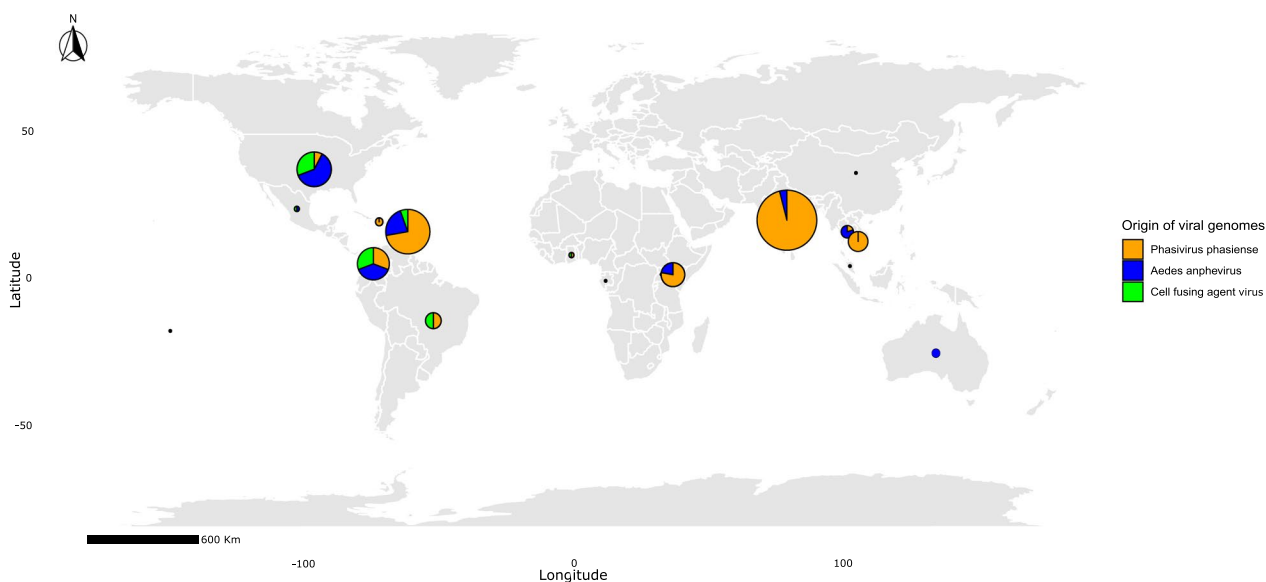
Overall, three distinct subgroups were identified for each virus on the basis of their geographic origin: Americas, Asia, and Africa (Fig. 2), revealing geographically structured dispersal patterns, whereas the AeAV sequence from Oceania exhibited closer affinity with Asian lineages (Fig. 2C). This suggests a recent introduction of these viruses (and consequently their mosquito vectors) into Oceania, likely originating from Asia. The analysis incorporated sequences from established mosquito colonies (Supplementary Table S1).

Another key observation is that the phylogenetic trees suggest that PCLV, CFAV, and AeAV populations in Africa and the Americas diverged from their Asian counterparts at an early point in their evolutionary history. However, as shown in Fig. 2, certain PCLV sequences from Asia and Africa appear to share a more recent common ancestor compared with those from the Americas and Africa. In particular, the Asian sequence MF614132.1 and the African sequence LC496491.1 suggest possible introductions of PCLV-infected vectors from Africa into Asia.

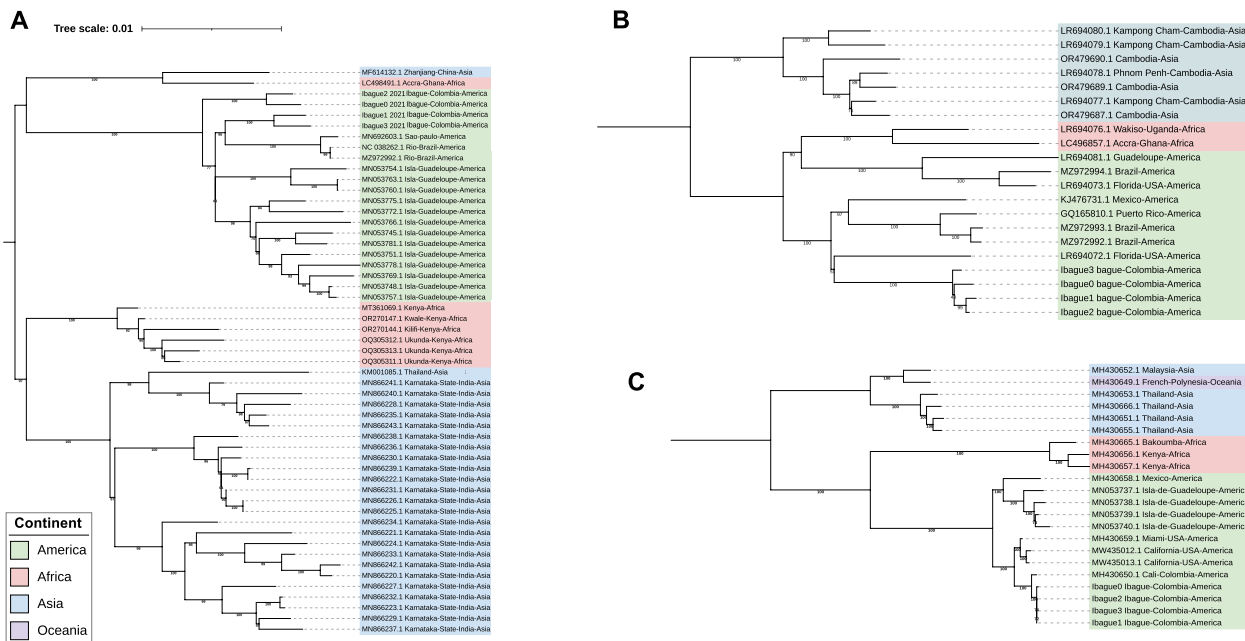
In addition, differences in the length of terminal branches were observed for all three viruses, even at local scales (i.e., sequences from the same geographic location). For example, in Fig. 2A, B, variation is evident among sequences collected in Ibagué in 2021.

### Genetic variability of AeAV, CFAV, and PCLV

To better understand the genetic variability of the studied viruses and their potential relationship with patterns of geographic dispersal, a genetic diversity analysis was conducted. This included the estimation of parameters



**Fig. 1** Map showing the geographical distribution of insect-specific viruses (ISVs), color-coded as follows: orange for *Phasivirus phasiense* (PCLV), blue for *Aedes anophevirus* (AeAV), and green for cell-fusing agent virus (CFAV). The sequences were obtained from vectors reported worldwide, with a total of 45 sequences from the Americas, 38 from Asia, 12 from Africa, and 1 from Australia. The size of each pie chart reflects the number of sequences obtained from each geographic region



**Fig. 2** Maximum likelihood phylogenetic trees of *Aedes aegypti* insect-specific viruses (ISVs) inferred using 1000 bootstrap replicates (values > 70% shown at nodes). **A** represents *Phasivirus phasiense* (PCLV), **B** shows cell-fusing agent virus (CFAV), and **C** displays *Aedes anophevirus* (AeAV). Sequence identifiers correspond to GenBank accession numbers, with tips color-coded by geographic origin: green (America), blue (Asia), red (Africa), and yellow (Oceania). Samples derived from laboratory colonies are marked with a black star (★)

such as nucleotide diversity, number of haplotypes, and the number of polymorphic sites, which allowed for the evaluation of the degree of differentiation among viral

populations (Table 1). Intraspecific genetic variability for each virus (with  $n > 1$  per continent) is detailed in Supplementary Table S2.

**Table 1** Genetic diversity of evaluated ISVs

Parameter	<i>Aedes anphevirus</i> (AeAV)	Cell-fusing agent virus (CFAV)	Phasivirus phasiense (PCLV)
Number of sequences	22	21	53
Genome length (bp)	12207	4948	5796
Polymorphic sites	1871	864	912
Nucleotide diversity ( $\pi \pm$ SD)	0.0470 $\pm$ 0.005	0.04217 $\pm$ 0.001	0.0258 $\pm$ 0.0006
Haplotypes (diversity)	21 (0.996)	21 (1.0)	50 (0.998)
$\Theta_w$ (per site $\pm$ SD)	0.04205 $\pm$ 0.013	0.04853 $\pm$ 0.016	0.03501 $\pm$ 0.009
$\Theta_w$ (per seq $\pm$ SD)	513.25 $\pm$ 165.84	240.111 $\pm$ 84.7	200.968 $\pm$ 56.2
Gaps/missing data	92	3	56

The table summarizes various genetic and diversity parameters for three viral species: *Aedes anphevirus* (AeAV), cell-fusing agent virus (CFAV), and Phasivirus phasiense (PCLV). The standard deviation to  $\Theta_w$  per site and sequence is standard deviation of theta (no recombination)

AeAV exhibited high genetic diversity, with a nucleotide diversity ( $\pi$ ) of 0.0470 and a Watterson’s Theta per site ( $\Theta_w$ ) of 0.04205—closely aligned values that suggest a scenario of neutral evolution. A total of 1871 polymorphic sites were identified within a 12,207 bp region, reflecting substantial genetic variability. Haplotype diversity ( $H_d$ ) was 0.996, with 21 haplotypes across 22 sequences, indicating that nearly all sequences were unique. The standard deviation (SD) of  $\pi$  was low (0.005), supporting the robustness of this estimate. The presence of 92 sites with gaps or missing data is not unexpected, given the nature of RNA viruses (Table 1).

Similarly, CFAV showed the highest genetic diversity among the viruses analyzed, with a  $\Theta_w$  per site of 0.04853—the highest in the study. Nucleotide diversity ( $\pi=0.04217$ ) was comparable to that of AeAV, but haplotype diversity reached the maximum value ( $H_d=1.0$ ), as all 21 analyzed sequences corresponded to distinct haplotypes. Only three sites contained gaps, indicating a high-quality alignment. The low  $\pi$  SD (0.005) and the relatively high SD in  $\Theta_w$  (0.016) suggest that while overall diversity is high, there may be subpopulations with distinct evolutionary dynamics (Table 1).

In this context, PCLV exhibited moderate diversity ( $\pi=0.0258$ ), lower than AeAV and CFAV, but with a high haplotype diversity ( $H_d=0.998$ ), as 50 out of 53 sequences represented unique haplotypes. The  $\Theta_w$  per site (0.03501) was higher than  $\pi$ , which could indicate purifying selection or a recent population expansion. A total of 912 polymorphic sites were found in a 5796-bp region, along with 56 sites containing gaps. Although the low  $\pi$  SD (0.004) suggests consistent results, the presence of missing data indicates that subsequent analyses should be interpreted with caution (Table 1).

Intraspecific genetic variability for each virus (with  $n > 1$  per continent) is detailed in Supplementary Table S2. To explore these patterns in greater detail, we performed region-specific analyses for the Americas,

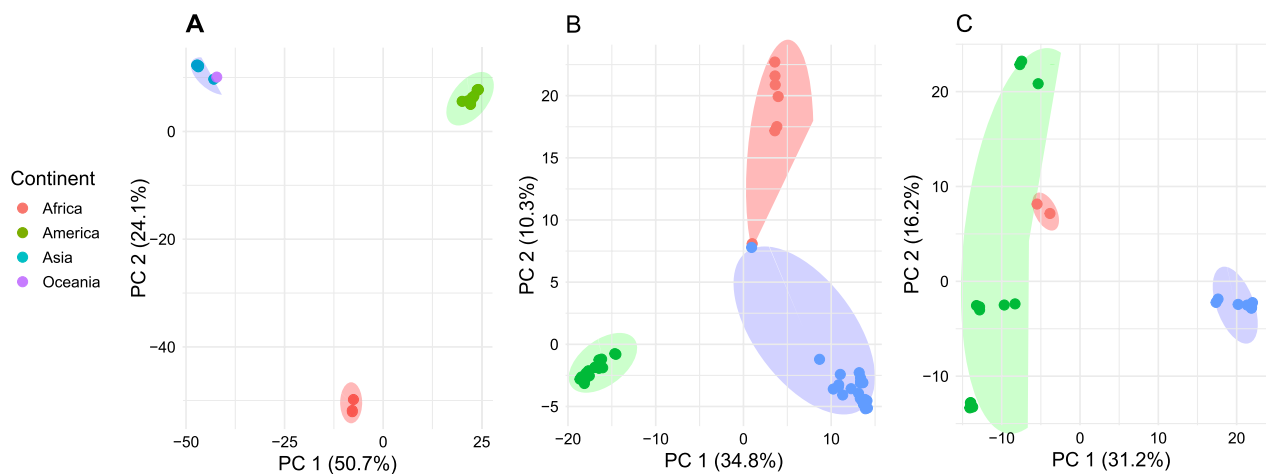
Asia, and Africa (Supplementary Table S2). These disaggregated results confirmed that AeAV exhibited its highest diversity in Asia ( $\pi=0.0443$ ), while CFAV was most diverse in the Americas ( $\pi=0.03469$ ). Haplotype diversity remained high across all viruses and regions, suggesting substantial intraregional variation.

#### Geographic pattern of genetic variability

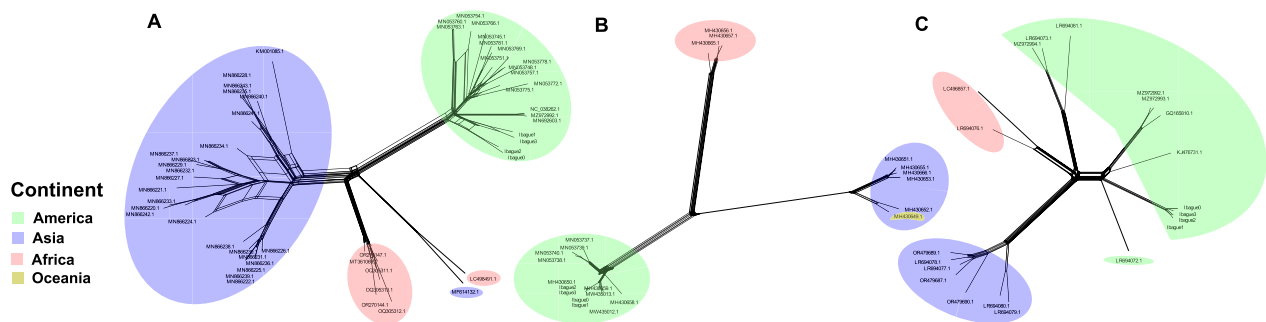
High values of nucleotide and haplotype diversity reflect divergent evolutionary dynamics among the analyzed ISVs. To assess whether this genetic variability is structured by geographic origin, a principal component analysis (PCA) was performed, which revealed distinct viral groupings by continent (Fig. 3A–C). Neotropical populations showed marked genetic differentiation compared with Asian populations, while African populations exhibited partial overlap with other geographic groups (Fig. 3B, C). This pattern was particularly evident for PCLV, which displayed overlap between African and Asian samples. American populations exhibited greater genetic variability, with CFAV sequences showing variable positions along the second principal component ( $y$ -axis), including overlap with a single African sequence (Fig. 3C). In line with the PCA, the phylogenomic analysis also identified clear geographic structuring among the viral lineages (Fig. 2A), and the PCA results show a clear separation of continental clusters, consistent with limited genetic exchange between these regions.

#### Exploration of other evolutionary events

The phylogenetic analyses conducted previously revealed the presence of three distinct lineages for each of the viruses studied. However, to avoid relying solely on strictly bifurcating methods such as phylogenetic trees and to explore potential recombination events, a phylogenetic network analysis was performed using SplitsTree version 4.19.2 (Fig. 4).



**Fig. 3** Principal component analysis (PCA) of *Aedes aegypti* insect-specific viruses: **A** *Aedes anophevirus* (AeAV), **B** *Phasivirus phasiense* (PCLV), and **C** cell-fusing agent virus (CFAV). Individual viral genomes are represented by points, and continental groups are indicated by ellipses. Colors indicate geographic origin—green (America), blue (Asia), red (Africa), yellow (Oceania)—with points and ellipses sharing the same color but differing in opacity to distinguish individuals from group clusters. Distinct clustering by continent is observed, particularly among American, African, and Asian populations



**Fig. 4** Phylogenetic networks of *Aedes aegypti* insect-specific viruses: **A** *Phasivirus phasiense* (PCLV), **B** *Aedes anophevirus* (AeAV), and **C** cell-fusing agent virus (CFAV). Sequences are color-coded by geographic origin: green (America), blue (Asia), yellow (Oceania), and red (Africa). The networks reveal: (1) distinct continental clustering (particularly America/Africa/Asia), suggesting phylogeographic structure; and (2) multiple reticulations in PCLV (**A**) and CFAV (**C**) networks, indicating genetic recombination events

The network analysis (SplitsTree, Fig. 4) revealed the clustering of sequences into three divergent groups corresponding to their geographic origin, consistent with the results of the phylogenomic and PCA analyses. In both PCLV and CFAV, multiple phylogenetic reticulations were observed in the network, indicating ongoing genetic recombination processes. This was statistically supported for PCLV by the phi test ( $P=2.116 \times 10^{-6}$ ), while for CFAV the phi test result ( $P=0.09765$ ) did not reach significance, indicating only a weak or nonsignificant signal of recombination.

In contrast, AeAV exhibited a network with few reticulations ( $P=0.9641$ ), suggesting an absence of recombination. Specific relationships were identified between African and Asian PCLV sequences LC498491.1 from Ghana and MF614132.1 from China.

The RDP analysis detected recombination events in PCLV, which were mainly found among sequences from the same geographic origin, particularly in regions of the L segment (2821–5418 bp) and the RdRp gene (1572–3792 bp) (Supplementary Fig. S1).

#### Estimation of TMRCA and temporal analysis of ISV spread in *Aedes aegypti*

On the basis of previous findings, we analyzed the temporal spread and time to the most recent common ancestor (TMRCA) of the ISVs CFAV and AeAV using TreeTime under molecular clock models. PCLV was excluded from this analysis, as recombination events in this virus occurred exclusively among sequences from the same geographic region, which may not reflect large-scale dispersal patterns. A similar number of available

sequences for AeAV and CFAV enabled a balanced comparison between models.

Time-scaled analysis using TimeTree:

To estimate divergence times and explore the evolutionary history of ISVs in *Ae. aegypti*, time-calibrated phylogenetic trees were constructed using BEAST1.

The time-calibrated phylogeny of AeAV revealed a clear geographic structure with well-defined separation between African, American, and Asian lineages (Fig. 5A). The analysis indicated an initial divergence event between Asian lineages and the African–American clade approximately 611 years ago ( $\pm 194.9$ ), followed by the separation of African and American lineages around 504 years ago ( $\pm 90.8$ ). Within the Asian lineage, sequences from Oceania diverged from a Malaysian lineage approximately 76 ( $\pm 14.6$ ) years ago, with Malaysian sequences showing the closest genetic affinity to Oceania strains.

The TimeTree analysis of CFAV revealed a distinct evolutionary pattern (Fig. 5B). Asian sequences diverged from the African and American lineages approximately 236 years ago ( $\pm 60.9$ ). Subsequently, around 203 years ago, most American sequences and all African sequences diverged. Finally, approximately 192 years ago ( $\pm 109.7$ ), a clear divergence event occurred where the American sequences (including those from Guadeloupe, Florida USA, and Brazil) completely separated from the African lineage, establishing distinct American and African clades.

## Discussion

Multiple studies have characterized the *Ae. aegypti* virome across diverse geographic regions, consistently identifying a core set of insect-specific viruses (ISVs), notably CFAV, AeAV, and PCLV [60, 61]. In parallel, recent research has explored the factors shaping the global distribution of *Ae. aegypti*, emphasizing the epidemiological relevance of understanding both natural and anthropogenic drivers of its dispersal [2, 62–64]. Herein, we integrated virome and empiric vector dispersal, providing evidence that ISVs associated with *Ae. aegypti* exhibit genetic diversity and empirical geographic structure that provides insights into the mosquito's dispersal dynamics at a continental scale [12].

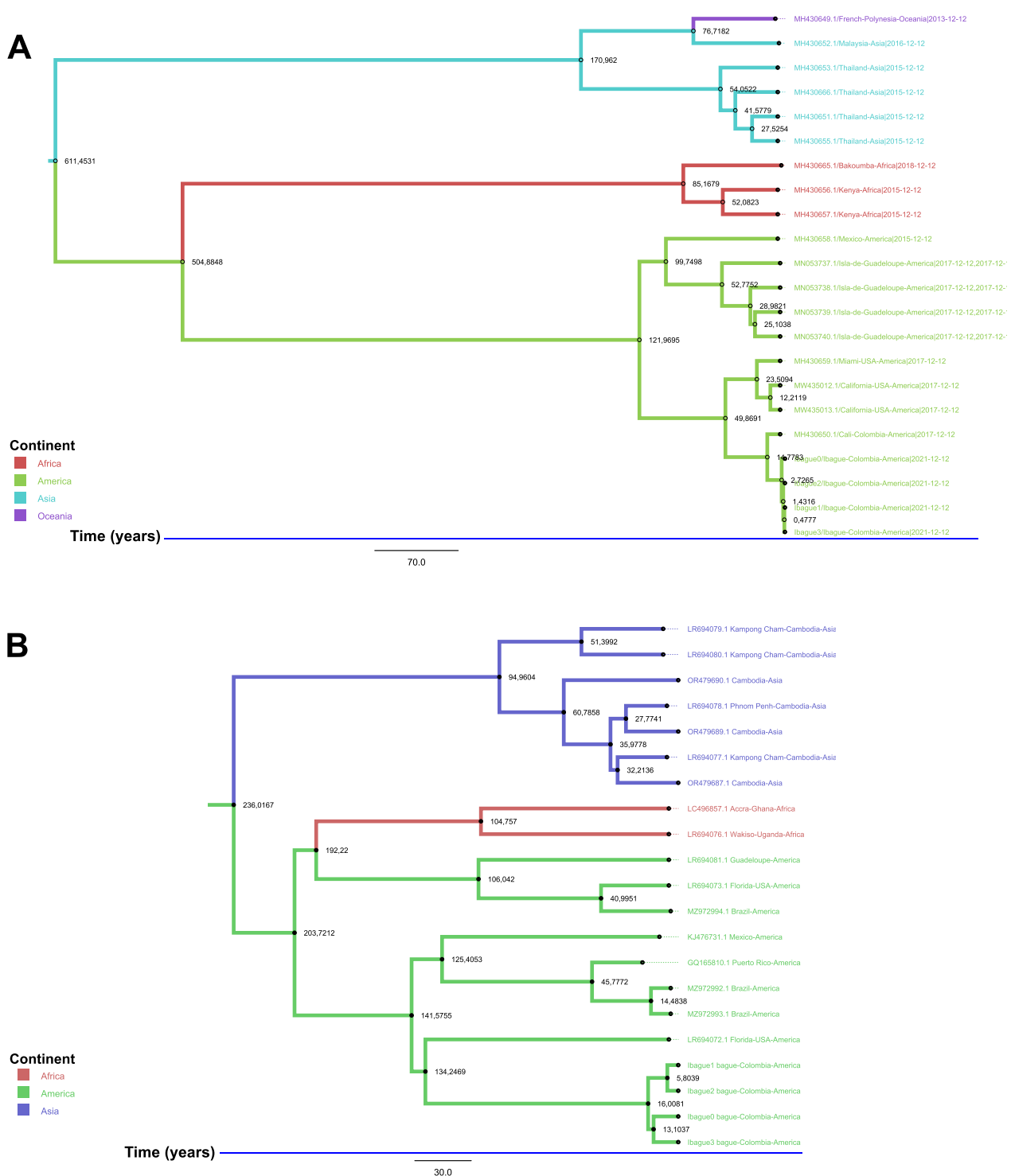
Our comprehensive analysis of three insect-specific viruses (ISVs)—CFAV, PCLV, and AeAV—revealed divergent evolutionary patterns. Recombination events were detected in PCLV but not in CFAV and AeAV. All three viruses exhibited high haplotype diversity, highlighting their evolutionary dynamism and potential as molecular markers of *Ae. aegypti* dispersal. As ISVs are exclusive to insect hosts and not influenced by vertebrate immune pressures or zoonotic cycles, they exhibit greater evolutionary stability and lower clearance rates [12, 16, 17, 65,

66]. Their long-term coevolution with *Ae. aegypti* positions them as potential proxies for investigating historical patterns of mosquito spread [67, 68]. Future efforts deepen the use of ISVs as phylogeographic tools, potentially incorporating additional viral taxa to strengthen inferences of vector movement across time and space.

Among the three viruses, CFAV demonstrated the highest nucleotide diversity ( $\pi=0.0421$ ) and haplotype diversity ( $Hd=1.0$ ; Table 1). Despite the presence of reticulations in the phylogenetic network, the phi test result ( $P=0.09765$ ) did not reach significance, indicating no strong evidence of recombination. These results build on previous work by Baidaliuk et al. [69] but differ in that the present dataset included sequences obtained directly from wild mosquitoes as well as from long-established laboratory colonies. While this approach avoids biases from extensive in vitro cell passages, we acknowledge that long-term laboratory maintenance can introduce its own artifacts, such as genetic drift or adaptation, and therefore, interpreted data from these colonies with caution. Phylogenetic and temporal analyses indicated the divergence of Asian lineages from the African–American cluster approximately 236 years ago (around the year 1785), followed by a separation between African and American lineages around 192 years ago ( $\sim 1829$ ; Fig. 5B). These estimates correspond to the late colonial and early post-colonial periods, when human-mediated trade and migration between continents intensified [2, 68]. CFAV thus holds promise as a molecular chronometer to trace historical *Ae. aegypti* movements. Expanding datasets with ancient samples or museum specimens could further refine these reconstructions.

AeAV showed no evidence of recombination (phi test  $P=0.9641$ ), yet exhibited the highest nucleotide diversity ( $\pi=0.0470$ ) and near-maximal haplotype diversity ( $Hd=0.996$ ; Table 1). Phylogenetic analysis revealed clear geographic structuring, with distinct African, Asian, and American clades (Fig. 2C), consistent with lineage diversification following geographic isolation. Temporal phylogenetic estimates revealed a deep divergence between African and Asian lineages approximately 611 years ago ( $\sim 1410$ ), and a subsequent split between African and American lineages  $\sim 504$  years ago ( $\sim 1517$ ; Fig. 5A). These timings coincide with the early phases of European colonial expansion and the onset of the transatlantic slave trade [2, 70], processes historically implicated in the spread of *Ae. aegypti* across continents [5, 71]. The alignment between AeAV divergence and historical human-driven dispersal supports the potential of ISVs as robust phylogeographic markers for reconstructing long-term mosquito movements.

Notably, the divergence times estimated from our ISV phylogenies present a different chronological order



**Fig. 5** Time-calibrated trees (TimeTree) for *Aedes anophevirus* (AeAV) (**A**) and cell-fusing agent virus (CFAV) (**B**), generated using BEAST1. A relaxed molecular clock was applied for sequence analysis, while a strict clock was used for time scaling and geographic assignment. Model convergence was validated by ensuring an effective sample size (ESS) greater than 200, guaranteeing the stability of the inferred parameters. The phylogenetic topology was reconstructed using maximum likelihood methods, providing a reliable representation of evolutionary relationships. In addition,  $1 \times 10^9$  MCMC (Markov chain Monte Carlo) iterations were implemented to obtain precise estimates of divergence times and the evolutionary dynamics of both viruses

compared with the established dispersal history of *Ae. aegypti*, which is based on the vector's own genomic markers. While mosquito genomics indicates a dispersal from Africa to the Americas followed by introductions into Asia [5, 8, 72], our viral data suggest an earlier divergence of the Asian lineages. We do not interpret this discrepancy as a contradiction but rather as evidence that ISVs provide a complementary layer of information. It is plausible that the evolutionary history of these viruses is not only shaped by the deep, historical migrations of their host but also by more recent or transient contact events between mosquito populations that did not result in significant host gene flow [73]. Therefore, ISV phylogeography may not simply recapitulate the vector's history but rather enriches it, offering insights into ecological and epidemiological links that are not apparent from host genetics alone.

A key advantage of using ISVs as phylogeographic markers is that their evolutionary dynamics are distinct from those of their mosquito host. With potentially faster mutation rates and different selection pressures [12, 16, 17, 65, 66], ISVs can act as high-resolution trackers for recent or transient events, such as inter-population contacts that may not result in lasting host gene flow. This is precisely how they complement the information from the mosquito genome: While the host's nuclear DNA reveals deep, long-term migration and isolation history [5, 8, 72], ISVs provide a more dynamic ledger of ecological and epidemiological connections. Our findings, which show a different historical narrative for the viruses compared with the vector, underscore this complementary role.

PCLV presented the lowest nucleotide diversity ( $\pi=0.0256$ ) among the three viruses, although haplotype diversity remained high ( $Hd=0.998$ ; Table 1). Recombination was detected within geographically restricted lineages (Supplementary Fig. S1), especially in the L segment (positions 2821–5418 and 1572–3792), consistent with modular evolution observed in segmented viruses [74]. The phi test confirmed significant recombination ( $P=4.162 \times 10^{-5}$ ). These patterns support the hypothesis that PCLV may be more suitable for studying fine-scale dispersal dynamics rather than broader transcontinental movements [75]. However, the exclusion of the M and S genome segments—often more variable—may underestimate total viral diversity [76]. Full genome sequencing, especially of the underexplored M and S segments, could reveal cryptic diversity in PCLV and improve its utility in vector tracking at multiple spatial scales.

Differences in the evolutionary trajectories of these ISVs may be attributed to intrinsic viral traits such as replication rate, interactions with the mosquito immune system, and vertical transmission modes [77,

78]. In addition, the degree of virus–vector coadaptation likely influences their suitability as molecular markers. Viruses such as AeAV, with deeper divergence times, may better reflect ancient mosquito dispersal, while CFAV and PCLV appear more informative of recent or regional movements (Hollingsworth et al. [12]). Dissecting the molecular interactions between ISVs and *Ae. aegypti* will provide deeper insight into their coevolution and potential as stable evolutionary indicators.

Our findings indicate that ISVs, owing to their potentially faster evolutionary rates, may capture recent and fine-scale population events that are often invisible to host-based genomic markers. For instance, the recombination detected within geographically restricted PCLV lineages is indicative of ongoing, localized viral evolution. Similarly, the more recent divergence estimates for CFAV, corresponding to the post-colonial era, suggest its potential utility in tracking mosquito dynamics on shorter time frames. Collectively, this suggests that ISVs may function not only as proxies for ancient history but also as potential high-resolution tools for understanding contemporary vector dispersal.

Our phylogenetic findings resonate with prior studies: Parry and Asgari [22] identified three monophyletic AeAV lineages corresponding to Africa, the Asia-Pacific region, and the Americas, consistent with our results and further supported by the single Oceanian sequence, which clustered within the Asian clade. Zhou et al. [79] observed shared ancestry between African and American CFAV strains, in line with our observations. Lole et al. [75] noted PCLV phylogeographic associations such as Ghana–China and USA–UK, also aligning with our results. These similarities bolster the robustness of our analyses and emphasize the reproducibility of ISV phylogeography across studies. Cross-validation of ISV phylogenies with mosquito population genomics could further strengthen inferences of vector movement and evolutionary history.

The timing of CFAV's diversification in the New World, with divergence events estimated at ~1785 and ~1829 (236 and 192 years ago, respectively; Fig. 5B), coincides with the late colonial and early post-colonial periods, when intensified trade and human mobility linked Africa and the Americas [2]. This reinforces its value as a marker for recent mosquito dispersal. In contrast, AeAV's elevated variability and deeper divergence times (~1410 and ~1517; Fig. 5A) suggest that it may better capture historical mosquito movements associated with early transoceanic exchanges. Although PCLV's segmented genome poses interpretative challenges, its localized recombination signatures offer unique insights into microevolutionary processes.

Integrating viral and vector genomic data with historical and ecological records can enhance the accuracy of dispersal modeling in vector-borne disease systems.

A major strength of this study lies in the reliance on sequences obtained directly from wild-caught mosquitoes, complemented by data from long-established laboratory colonies originally derived from field populations, which minimizes potential laboratory-induced biases [79]. By employing a comparative approach across ISVs with varying genome architectures and evolutionary rates, we demonstrate the potential of combining phylogenomics, recombination analysis, and molecular-clock dating to track vector dynamics. To our knowledge, this is among the first studies to holistically assess multiple ISVs as indirect indicators of global mosquito mobility. This approach may serve as a blueprint for future vector studies seeking to integrate molecular epidemiology with evolutionary ecology.

Nonetheless, some limitations must be acknowledged. Geographic sampling was uneven (Americas: 45 sequences; Asia: 38; Africa: 13; Oceania: 3; Fig. 1), which may skew regional inferences and limit the resolution of phylogeographic patterns. The dataset is particularly biased by the underrepresentation of Africa and Oceania—two regions that were central to the historical dispersal of *Ae. aegypti*. This limitation must be taken into account when interpreting phylogenetic splits involving these regions, as it could underestimate their contribution to global virus diversity. Partial genome representation (e.g., only the L segment of PCLV) may also mask additional diversity or recombination signals [76]. While PCLV exhibited clear evidence of recombination, CFAV showed only weak signals and AeAV none, underscoring the need for broader genomic coverage and comparative analyses. The absence of prior studies on recombination in AeAV and PCLV further restricts contextual interpretation. Recombination may act as a compensatory mechanism in RNA viruses to mitigate deleterious mutations—a hypothesis supported by the “Müller’s ratchet” model [80]. Future work should prioritize balanced sampling from underrepresented regions, along with full-genome analyses and functional assays of recombination, to better resolve the evolutionary forces shaping ISV diversity.

Finally, these findings carry important implications for vector surveillance and arbovirus risk prediction. ISVs may reflect colonization history and mobility of *Ae. aegypti*, offering early warning proxies for arbovirus outbreaks. Principal component analysis (PCA) revealed clear continental clustering (Fig. 3), particularly in the Americas and Asia, while African sequences were more diffuse—potentially due to complex

migration histories or gene flow. By mapping ISV phylogeography, we can refine vector control strategies, especially in regions undergoing rapid urbanization or environmental change. Future studies should aim to integrate these molecular insights with real-time vector surveillance to predict and preempt emerging public health threats.

A promising perspective lies in the isolation and experimental propagation of ISVs, which would enable controlled studies to directly estimate mutation rates, recombination frequencies, and adaptive dynamics—parameters currently inferred indirectly from metagenomic datasets [81]. While metagenomic studies offer broad insights into viral diversity and distribution, they often rely on consensus sequences and pooled samples, potentially underestimating intrahost variability and transient viral subpopulations. In contrast, isolating ISVs and passaging them under defined conditions (e.g., in cell lines or live mosquitoes) would allow precise measurement of evolutionary rates, identification of mutational hotspots, and testing of selective pressures in real time [82]. This approach could clarify whether observed phylogenetic divergence reflects true mutation accumulation or artifacts of sequencing and assembly pipelines [83]. Moreover, comparing mutation rates derived from isolation studies with those inferred from metagenomics would help validate molecular clock models and refine the use of ISVs as tracers of mosquito population history [12].

## Conclusions

Our findings highlight the utility of insect-specific viruses (ISVs) as informative molecular markers that complement and enrich our understanding of the evolutionary and dispersal history of *Ae. aegypti*. Our multi-virus approach reveals a multilayered historical narrative: AeAV provides a window into ancient dispersal events linked to early colonial exchanges, while CFAV tracks more recent dynamics from the post-colonial era, and PCLV illuminates localized, micro-evolutionary processes. This demonstrates that it is the integration of these complementary viral histories that allows for a more holistic reconstruction of vector movement across different spatial and temporal scales. Beyond methodological innovation, this virome-based framework offers a novel lens for monitoring mosquito mobility and anticipating future patterns of vector spread in an era of accelerating climate change and global connectivity. Ultimately, harnessing the information encoded in the virome represents a powerful, next-generation strategy for the surveillance and control of one of the world’s most important disease vectors.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13071-025-07120-3>.

Additional file 1. Fig. S1. Recombination analysis of *Phasivirus phasiense* (PCLV).

Additional file 2. Table S1. Geographic distribution and genetic data of insect-specific viruses (ISVs) in *Ae. aegypti*.

Additional file 3. Table S2. Genetic diversity parameters of ISVs by continent.

### Acknowledgements

We thank the Laboratorio de Investigaciones en Parasitología Tropical (LIPT) at Universidad del Tolima for their pivotal role in mosquito collection, sequencing, viral assembly, and taxonomic assignments from Ibagué and Juan David Rojas for his collaboration in these activities.

### Author contributions

Juan David Ramírez: conceptualization, supervision, funding acquisition, and writing—review and editing. Weimar D. Briñez: conceptualization, methodology, formal analysis, and writing—original draft. Marina Muñoz and Luz H. Patiño: methodology, bioinformatics analysis, and validation. Daniel A. Urrea: data curation (LIPT activities) and resources. All authors contributed to manuscript revision and approved the final version.

### Funding

This study was funded by Vicerrectoría de Investigaciones from Universidad del Rosario and Universidad del Tolima (Vice-Rectorate for Research) for mosquito collection, sequencing, viral assembly, taxonomic assignments, and computational infrastructure at LIPT.

### Availability of data and materials

Sequence data that support the findings of this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession no. PRJNA1257366.

### Declarations

#### Ethical approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Centro de Investigaciones en Microbiología y Biotecnología-UR (CIMBIUR), School of Sciences and Engineering, Universidad del Rosario, Bogotá, Colombia. <sup>2</sup>Laboratorio de Investigaciones en Parasitología Tropical (LIPT), Facultad de Ciencias, Universidad del Tolima, Ibagué, Colombia. <sup>3</sup>Instituto de Biotecnología-UN (IBUN), Universidad Nacional de Colombia, Bogotá, Colombia. <sup>4</sup>Center for Global Health and Interdisciplinary Research, Department of Global, Environmental and Genomic Health Sciences, College of Public Health, University of South Florida, Tampa, FL, USA.

Received: 5 May 2025 Accepted: 19 October 2025

Published online: 18 November 2025

### References

- Kröpelin S, Verschuren D, Lézine AM, Eggermont H, Cocquyt C, Francus P, et al. Climate-driven ecosystem succession in the Sahara: the past 6000 years. *Science*. 2008;320:765–8.

- Powell JR, Tabachnick WJ. History of domestication and spread of *Aedes aegypti*- A review. *Mem Inst Oswaldo Cruz*. 2013;108:11–7.
- Brown JE, Evans BR, Zheng W, Obas V, Barrera-Martinez L, Egizi A, et al. Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. *Evolution*. 2014;68:514–25.
- Rose NH, Badolo A, Sylla M, Akorli J, Otoo S, Gloria-Soria A, et al. Dating the origin and spread of specialization on human hosts in *Aedes aegypti* mosquitoes. *Elife*. 2023;12:e83524.
- Soghigian J, Gloria-Soria A, Robert V, Le Goff G, Failloux AB, Powell JR. Genetic evidence for the origin of *Aedes aegypti*, the yellow fever mosquito, in the southwestern Indian Ocean. *Mol Ecol*. 2020;29:3593–606.
- Lee Y, Schmidt H, Collier TC, Conner WR, Hanemaaijer MJ, Slatkin M, et al. Genome-wide divergence among invasive populations of *Aedes aegypti* in California. *BMC Genomics*. 2019;20:204.
- Hopperstad KA, Reiskind MH, Labadie PE, Burford Reiskind MO. Patterns of genetic divergence among populations of *Aedes aegypti* L. (Diptera: Culicidae) in the southeastern USA. *Parasit Vectors*. 2019;12:511.
- Gloria-Soria A, Ayala D, Bheecarry A, Calderon-Arguedas O, Chadee DD, Chiappero M, et al. Global genetic diversity of *Aedes aegypti*. *Mol Ecol*. 2016;25:5377–95.
- Moore M, Sylla M, Goss L, Burugu MW, Sang R, Kamau LW, et al. Dual African origins of global *Aedes aegypti* s.l. populations revealed by mitochondrial DNA. *PLoS Negl Trop Dis*. 2013;7:e2175.
- Maynard AJ, Ambrose L, Bangs MJ, Ahmad R, Butafa C, Beebe NW. Population structure and invasion history of *Aedes aegypti* (Diptera: Culicidae) in Southeast Asia and Australasia. *Evol Appl*. 2023;16:849–62.
- Monsalve Y, Triana-Chávez O, Gómez-Palacio A. Population structure and ancestry prediction of *Aedes aegypti* (Diptera: Culicidae) supports a single African origin of Colombian populations. *Mem Inst Oswaldo Cruz*. 2021;116:e200441.
- Hollingsworth BD, Grubaugh ND, Lazzaro BP, Murdock CC. Leveraging insect-specific viruses to elucidate mosquito population structure and dynamics. *PLoS Pathog*. 2023;19:e1011588.
- Martínez D, Gómez M, Hernández C, Muñoz M, Campo-Palacio S, González-Robayo M, et al. Emergence of dengue virus serotype 2 cosmopolitan genotype, Colombia. *Emerg Infect Dis*. 2024;30:5.
- Logan RAE, Quek S, Muthoni JN, von Eicken A, Brettell LE, Anderson ER, et al. Vertical and horizontal transmission of cell fusing agent virus in *Aedes aegypti*. *Appl Environ Microbiol*. 2022;88:e01062-22.
- Moonen JP, Schinkel M, van der Most T, Miesen P, van Rij RP. Composition and global distribution of the mosquito virome - a comprehensive database of insect-specific viruses. *One Health*. 2023;16:100490.
- Suzuki Y, Suzuki T, Miura F, Reyes JIL, Asin ICA, Mitsunari W, et al. No detectable fitness cost of infection by cell-fusing agent virus in *Aedes aegypti* mosquitoes. *R Soc Open Sci*. 2024;11:231373.
- Bolling BG, Weaver SC, Tesh RB, Vasilakis N. Insect-specific virus discovery: significance for the arbovirus community. *Viruses*. 2015;7:4911–28.
- Gómez M, Martínez D, Páez-Triana L, Luna N, De las salas JL, Hernández C, et al. Characterizing viral species in mosquitoes (Culicidae) in the Colombian Orinoco: insights from a preliminary metagenomic study. *Sci Rep*. 2023;13:22081.
- Zakrzewski M, Rašić G, Darbro J, Krause L, Poo YS, Filipović I, et al. Mapping the virome in wild-caught *Aedes aegypti* from Cairns and Bangkok. *Sci Rep*. 2018;8:4690.
- Olmo RP, Todjro YMH, Aguiar ERGR, de Almeida JPP, Ferreira FV, Armache JN, et al. Mosquito vector competence for dengue is modulated by insect-specific viruses. *Nat Microbiol*. 2023;8:135–49.
- Ripoll L, Iserte J, Cerrudo CS, Presti D, Serrat JH, Poma R, et al. Insect-specific RNA viruses detection in field-caught *Aedes aegypti* mosquitoes from Argentina using NGS technology. *PLoS Negl Trop Dis*. 2025;19:e0012792.
- Parry R, Asgari S. *Aedesanphevirus*: an insect-specific virus distributed worldwide in *Aedes aegypti* mosquitoes that has complex interplays with *Wolbachia* and dengue virus infection in cells. *J Virol*. 2018;92:e00224-18.
- de Almeida JP, Aguiar ER, Armache JN, Olmo RP, Marques JT. The virome of vector mosquitoes. *Curr Opin Virol*. 2021;49:7–12.
- Calle-Tobón A, Pérez-Pérez J, Forero-Pineda N, Chávez OT, Rojas-Montoya W, Rúa-Urbe G, et al. Local-scale virome depiction in Medellín, Colombia, supports significant differences between *Aedes aegypti* and *Aedes albopictus*. *PLoS ONE*. 2022;17:e0263143.

25. Harvey E, Holmes EC. Diversity and evolution of the animal virome. *Nat Rev Microbiol.* 2022;20:321–34.
26. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, et al. Redefining the invertebrate RNA virosphere. *Nature.* 2016;540:539–43.
27. Sandoval-Pineda JF, Ochoa-Corona FM, Torres-Rojas E, Sandoval-Pineda JF, Ochoa-Corona FM, Torres-Rojas E. Evaluación de diferentes métodos de extracción de ARN a partir del hongo nativo *Xylaria* sp. *Rev Colomb Biotechnol.* 2017;19:42–54.
28. Portillo M, Fenoll C, Escobar C. Evaluation of different RNA extraction methods for small quantities of plant tissue: combined effects of reagent type and homogenization procedure on RNA quality-integrity and yield. *Physiol Plant.* 2006;128:1–7.
29. Wingett SW, Andrews S. FastQ screen: a tool for multi-genome mapping and quality control. *F1000Res.* 2018;7:1338.
30. University of Washington Bothell/STEM Linux/prinseq uwb-linux/prinseq. 2024. <https://github.com/uwb-linux/prinseq>. Accessed 22 Apr 2025.
31. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20.
32. Trapnell C, Pachter L, Salzberg SL. Tophat: discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25:1105–11.
33. Giraldo-Calderón GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, et al. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic Acids Res.* 2015;43:D707–713.
34. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol.* 2011;29:644–52.
35. Fass J. *count\_fasta.pl* [computer program]. David (CA): The Bioinformatics Core at UC Davis Genome Center; 2010.
36. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2022;50:D20–6.
37. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics.* 2009;25:1968–9.
38. Otto TD, Dillon GP, Degraeve WS, Berriman M. RATT: rapid annotation transfer tool. *Nucleic Acids Res.* 2011;39:e57.
39. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience.* 2021;10:giab008.
40. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics.* 2012;28:2678–9.
41. Aguiar ERGR, Olmo RP, Paro S, Ferreira FV, de Faria IJS, Todjro MYM, et al. Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. *Nucleic Acids Res.* 2015;43:6191–206.
42. Bolling BG, Vasilakis N, Guzman H, Widen SG, Wood TG, Popov VL, et al. Insect-specific viruses detected in laboratory mosquito colonies and their potential implications for experiments evaluating arbovirus vector competence. *Am J Trop Med Hyg.* 2015;92:422–8.
43. Chang P, Gao H, Sun Q, He X, Huang F. Retraction: transcriptome profiling reveals new insights into the roles of neuronal nitric oxide synthase on macrophage polarization towards classically activated phenotype. *PLoS ONE.* 2021;16:e0260912.
44. Xu H, Tong G, Yan T, Dong L, Yang X, Dou D, et al. Transcriptomic analysis provides insights to reveal the bmp6 function related to the development of intermuscular bones in zebrafish. *Front Cell Dev Biol.* 2022;10:821471. <https://doi.org/10.3389/fcell.2022.821471/full>.
45. Patzina-Mehling C, Kopp A, Rauhöft L, Şuleşço T, Jones TC, Drosten C, et al. Genomic surveillance indicates high site-specific heterogeneity of West Nile virus in mosquitoes in rural regions of Germany across seasons. *One Health.* 2025. <https://doi.org/10.1016/j.onehlt.2025.101179>.
46. Melchert J, Radbruch H, Hanitsch LG, Baylis SA, Beheim-Schwarzbach J, Bleicker T, et al. Whole genome sequencing reveals insights into hepatitis E virus genome diversity, and virus compartmentalization in chronic hepatitis E. *J Clin Virol.* 2023;168:105583.
47. Zhang S, Wang B, Wan L, Li LM. Estimating Phred scores of Illumina base calls by logistic regression and sparse modeling. *BMC Bioinform.* 2017;18:335.
48. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772–80.
49. Okonechnikov K, Golosova O, Fursov M, the UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012;28:1166–7.
50. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37:1530–4.
51. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14:587–9.
52. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 2024;52:W78–82.
53. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol.* 2017;34:3299–302.
54. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 2006;23:254–67.
55. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, et al. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* 2021;7:veaa087.
56. Frias-De-Diego A, Jara M, Pecoraro BM, Crisci E. Whole genome or single genes? A phylodynamic and bibliometric analysis of PRRSV. *Front Vet Sci.* 2021;8:658512.
57. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol.* 2019;15:e1006650.
58. Martínez D, Gómez M, Hernández C, Campo-Palacio S, González-Robayo M, Montilla M, et al. Cryptic transmission and novel introduction of dengue 1 and 2 genotypes in Colombia. *Virus Evol.* 2024;10:veae068.
59. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 2018;67:901–4.
60. Mbigha Donfack KC, De Coninck L, Ghogomu SM, Matthijnsens J. *Aedes* mosquito virome in southwestern Cameroon: lack of core virome, but a very rich and diverse virome in *Ae. africanus* compared to other *Aedes* species. *Viruses.* 2024;16:1172.
61. Coninck LD, Matthijnsens J. The mosquito core virome: beyond the buzz. *Trends Parasitol.* 2024;40:201–2.
62. Kent TV, Schrider DR, Matute DR. Demographic history, genetic load, and the efficacy of selection in the globally invasive mosquito *Aedes aegypti*. *Genome Biol Evol.* 2025;17:evaf066. <https://doi.org/10.1093/gbe/evaf066>.
63. Laporta GZ, Potter AM, Oliveira JFA, Bourke BP, Pecor DB, Linton YM. Global distribution of *Aedes aegypti* and *Aedes albopictus* in a climate change scenario of regional rivalry. *Insects.* 2023;14:49.
64. Powell JR, Gloria-Soria A, Kotsakiozi P. Recent history of *Aedes aegypti*: vector genomics and epidemiology records. *Bioscience.* 2018;68:854–60.
65. Bello-Rodríguez BM, Vega-Jiménez J, Cañete R, Rodríguez-Morales AJ. Emergence of Oropouche virus infection in Matanzas, Cuba, 2024. *J Infect.* 2025;90:106470.
66. Daudt-Lemos M, Ramos-Silva A, Faustino R, de Noronha TG, de Vianna RAO, Cabral-Castro MJ, et al. Rising incidence and spatiotemporal dynamics of emerging and reemerging arboviruses in Brazil. *Viruses.* 2025;17:158.
67. Coatsworth H, Bozic J, Carrillo J, Buckner EA, Rivers AR, Dinglasan RR, et al. Intrinsic variation in the vertically transmitted core virome of the mosquito *Aedes aegypti*. *Mol Ecol.* 2022;31:2545–61.
68. Lewis J, Gallichotte EN, Randall J, Glass A, Foy BD, Ebel GD, et al. Intrinsic factors driving mosquito vector competence and viral evolution: a review. *Front Cell Infect Microbiol.* 2023;21:1330600. <https://doi.org/10.3389/fcimb.2023.1330600/full>.
69. Baidaliuk A, Lequime S, Moltini-Conclois I, Dabo S, Dickson LB, Prot M, et al. Novel genome sequences of cell-fusing agent virus allow comparison of virus phylogeny with the genetic structure of *Aedes aegypti* populations. *Virus Evol.* 2020;6:veaa018.
70. Zhao M, Ran X, Bai Y, Ma Z, Gao J, Xing D, et al. Genetic diversity of *Aedes aegypti* and *Aedes albopictus* from cohabiting fields in Hainan Island and the Leizhou Peninsula, China. *Parasit Vectors.* 2023;16:319.

71. Smith CEG. The history of dengue in tropical Asia and its probable relationship to the mosquito *Aedes aegypti*. *J Trop Med Hyg.* 1956;59:243–51.
72. Crawford JE, Balcazar D, Redmond S, Rose NH, Youd HA, Lucas ER, et al. 1206 genomes reveal origin and movement of *Aedes aegypti* driving increased dengue risk. *Science.* 2025;389:eads3732.
73. Hassel C, Mirand A, Farkas A, Diedrich S, Huemer HP, Peigue-Lafeuille H, et al. Phylogeography of coxsackievirus A16 reveals global transmission pathways and recent emergence and spread of a recombinant genotype. *J Virol.* 2017;91:e010-1128.
74. Sun MH, Ji YF, Li GH, Shao JW, Chen RX, Gong HY, et al. Highly adaptive Phenuiviridae with biomedical importance in multiple fields. *J Med Virol.* 2022;94:2388–401.
75. Lole K, Ramdasi A, Patil S, Thakar S, Nath A, Ghuge O, et al. Abundance of Phasi-Charoen-like virus in *Aedes aegypti* mosquito populations in different states of India. *PLoS ONE.* 2022;17:e0277276.
76. Ramos-Nino ME, Fitzpatrick DM, Eckstrom KM, Tighe S, Hattaway LM, Hsueh AN, et al. Metagenomic analysis of *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes from Grenada, West Indies. *PLoS ONE.* 2020;15:e0231047.
77. Gabiane G, Yen P, Failloux A. *Aedes* mosquitoes in the emerging threat of urban yellow fever transmission. *Rev Med Virol.* 2022;32:e2333.
78. Morais P, Trovão NS, Abecasis AB, Parreira R. Insect-specific viruses in the Parvoviridae family: genetic lineage characterization and spatiotemporal dynamics of the recently established *Brevihamaparvovirus* genus. *Virus Res.* 2022;313:198728.
79. Zhou N, Huang E, Guo X, Xiong Y, Xie J, Cai T, et al. Cell fusing agent virus isolated from Aag2 cells does not vertically transmit in *Aedes aegypti* via artificial infection. *Parasit Vectors.* 2023;16:402.
80. Dolan PT, Whitfield ZJ, Andino R. Mechanisms and concepts in RNA virus population dynamics and evolution. *Annu Rev Virol.* 2018;5:69–92.
81. Sanjuán R, Domingo-Calap P. Mechanisms of viral mutation. *Cell Mol Life Sci.* 2016;73:4433–48.
82. Cobbin JC, Charon J, Harvey E, Holmes EC, Mahar JE. Current challenges to virus discovery by meta-transcriptomics. *Curr Opin Virol.* 2025;51:48–55.
83. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet.* 2008;9:267–76.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.